M 332 Units 1–3

THE OPEN UNIVERSITY
Complex Analysis
Mathematics : A Third Level Course

Unit 1 Complex Numbers
Unit 2 Continuous Functions
Unit 3 Differentiation

COMPLEX ANALYSIS

THE OPEN UNIVERSITY

Mathematics: A Third Level Course

Complex Analysis Units 1, 2 and 3

Complex Numbers
Continuous Functions
Differentiation

Prepared by the Course Team

The Open University Press

The Open University Press, Walton Hall, Milton Keynes.

First published 1974. Reprinted 1979, 1983, 1986.

Copyright © 1974 The Open University.

All rights reserved. No part of this work may be reproduced in any form, by mimeograph or any other means, without permission in writing from the publishers.

Produced in Great Britain by
Speedlith Photo Litho Limited, Longford Trading Estate, Manchester M32 0JT.

ISBN 0 335 05550 8

This text forms part of the correspondence element of an Open University Third Level Course.
The complete list of units in the course is given at the end of this text.

For general availability of supporting material referred to in this text, please write to: Open University Educational Enterprises, 12 Cofferidge Close, Stony Stratford, Milton Keynes MK11 1BY, Great Britain.

Further information on Open University courses may be obtained from The Admissions Office, The Open University, P.O. Box 48, Milton Keynes, MK7 6AB.

1.4

CONTENTS

Page

Unit 1 Complex Numbers

Conventions 6
1.0 **Introduction** 7
1.1 **The Complex Number System** 8
1.2 **Problems** 16
1.3 **The Complex Plane** 20
1.4 **Problems** 29
1.5 **Sets of Points in the Complex Plane** 36
1.6 **Problems** 42

Unit 2 Continuous Functions

Conventions 48
2.0 **Introduction** 49
2.1 **Functions** 50
2.2 **Problems** 57
2.3 **Limits, Continuity and Open Sets** 59
2.4 **Problems** 67
2.5 **Closed Sets** 69
2.6 **Problems** 75
2.7 **Connectedness** 77
2.8 **Problems** 86
2.9 **Compactness** 88
2.10 **Problems** 95

Unit 3 Differentiation

Conventions 100
3.0 **Introduction** 101
3.1 **Derivatives** 102
3.2 **Interpreting the Derivative** 109
3.3 **Problems** 115
3.4 **The Exponential, Trigonometric and Hyperbolic Functions** . 118
3.5 **Problems** 127
3.6 **Logarithms and Powers** 129
3.7 **Problems** 137
3.8 **The Cauchy–Riemann Equations** 139
3.9 **Problems** 150
Notation 153
Index 155

Unit 1 Complex Numbers

Conventions

Before working through this text make sure you have read *A Guide to the Course: Complex Analysis*.

References to units of other Open University courses in mathematics take the form:

Unit M100 13, Integration II.

The set book for the course M231, Analysis, is M. Spivak, *Calculus*, paperback edition (W. A. Benjamin/Addison-Wesley, 1973). This is referred to as:

Spivak.

1.0 INTRODUCTION

You have almost certainly met complex numbers before, but you may well not have had much experience in manipulating them. In this unit we provide you with an opportunity to gain confidence in working with complex numbers by working a number of suitable problems. It is important that you take this task seriously because your success in the remainder of the course will depend largely upon a firm foundation in the manipulative aspects of the complex number system.

Perhaps the most striking difference between the real numbers and the complex numbers is the fact that complex numbers have a two-dimensional character, arising from our definition of a complex number as an ordered pair of real numbers. This two-dimensional aspect of complex numbers leads to a most useful representation of them as points in the plane.

In later units we shall begin the analysis of complex functions and it will be necessary for us to introduce a measure of the "closeness" of two complex numbers. We know that the distance between points in a plane can be measured by the usual Euclidean measure of distance, and this leads us to the important modulus function for complex numbers. This function will clearly play an important role in complex analysis if the subject is to develop along lines resembling real analysis. Later we will see how the complex modulus can be used to generalize many of the limiting processes of real analysis.

You may recall that many proofs in real analysis use the concept of least upper bound. This concept depends in turn on the relation of inequality which is defined in terms of positive and negative numbers. We shall show in this unit that it is not possible to define such a thing as a "positive" complex number and so least upper bound arguments will not carry over to complex analysis.

In real analysis, the idea of a least upper bound is used to develop the method of proof by repeated bisection, whose validity rests on the Nested Intervals Theorem. Because complex numbers are defined as pairs of real numbers, we are able to generalize this theorem to a Nested Rectangles Theorem, which will play a similar role in complex analysis to that of the Nested Intervals Theorem (or, equivalently, the least upper bound axiom) of real analysis.

There are three reading sections in this unit, each of which is followed by a problems section. In Section 1.1, the complex number system is defined; in Section 1.3 the relationship between the complex numbers and points in the plane is developed; in Section 1.5 various useful subsets of the complex number system are defined and the Nested Rectangles Theorem is proved.

1.1 THE COMPLEX NUMBER SYSTEM

In this section we shall define the complex number system as the set $\mathbf{R} \times \mathbf{R}$ (the Cartesian product of the set of reals, $\mathbf{R}$, with itself) with suitable addition and multiplication operations. We shall define the real and imaginary parts of a complex number and compare the properties of the complex number system with those of the real number system, particularly from the point of view of analysis.

* * * * * * * *

In complex analysis we are concerned with functions whose domains and codomains are subsets of the set of complex numbers. As you probably know, this structure is obtained from the set $\mathbf{R} \times \mathbf{R}$ by defining suitable operations of addition and multiplication. This reveals immediately one important difference between real analysis and complex analysis: in real analysis we are concerned with sets of real numbers, in complex analysis we are concerned with sets of *ordered pairs* of real numbers.

If you have met complex numbers before it could have been in several different contexts. For example, in the Foundation Course they were introduced in a geometric context. If you have studied M202, Topics in Pure Mathematics, you may know that the complex numbers can be obtained as an extension field of the real numbers. Whatever context is used to introduce complex numbers, one sooner or later meets the symbol i and the strange formula $i^2 = -1$.

Historically, the notion of a "number" i with this property arose from the desire to extend the real number system so that equations such as $x^2 + 1 = 0$ have solutions. There is no *real* number satisfying this equation so, as usual in mathematics, it was decided to *invent* a number system that did contain a solution. The remarkable fact is that having invented a solution of this *one* equation, we can use it to construct a system that contains the solutions of *every* polynomial equation!

For example, using the well-known formulas

$$\frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

for the solutions of the equation

$$ax^2 + bx + c = 0, \quad \text{where } a \neq 0,$$

we find that the solutions of

$$x^2 + x + 1 = 0$$

are apparently given by the expressions

$$\frac{-1 + \sqrt{-3}}{2}, \quad \text{and} \quad \frac{-1 - \sqrt{-3}}{2}.$$

These make no sense at all until we turn a blind eye to $\sqrt{-3}$ and just manipulate it formally, as though we knew what we were doing, to give

$$\sqrt{-3} = \sqrt{3 \times -1} = \sqrt{3}\sqrt{-1}.$$

We then say that if $i^2 = -1$ then we might as well press on and replace $\sqrt{-1}$ by i and so the "solutions" of the equation are

$$\frac{-1}{2} + \frac{i\sqrt{3}}{2}, \quad \text{and} \quad \frac{-1}{2} - \frac{i\sqrt{3}}{2}.$$

We can check our formal manipulation of square root signs by substituting these "solutions" into the equation and see that they work. But before we can do that, we need to know how to "add" and "multiply" "numbers" of the form $x + iy$, where x and y are real. If we assume that they add and multiply according to the ordinary laws of algebra, in other words that addition and multiplication are associative and commutative and that multiplication is distributive over addition, then we have

$$(a + ib) + (c + id) = (a + c) + i(b + d)$$

and

$$(a + ib) \times (c + id) = ac + i(bc + ad) + i^2bd.$$

If we replace i^2 by -1, we obtain

$$(a + ib) \times (c + id) = (ac - bd) + i(bc + ad).$$

Now you can substitute $-\frac{1}{2} \pm \frac{i\sqrt{3}}{2}$ into $x^2 + x + 1$ and see that our formal juggling with square root signs appears to be justified.

In the same way, we can factorize formally the expression $x^2 + bx + c$ into

$$\left[x - \left(\frac{-b + \sqrt{b^2 - 4c}}{2}\right)\right]\left[x - \left(\frac{-b - \sqrt{b^2 - 4c}}{2}\right)\right],$$

and hence "solve" any quadratic equation by assuming also that the property $\alpha\beta = 0$ if and only if $\alpha = 0$ or $\beta = 0$, which does apply to **R**, carries over to our new system. (Note the use of "or" in the "or both" sense. See *Unit M100 11, Logic I*.)

It may disconcert you to see us play around with a symbol that behaves as though it were the square root of -1: it should because, as yet, we have given no formal definition of this symbol, i. The problem that Pythagoras had in relating $\sqrt{2}$ to the set of rationals is presumably similar to that which we have in relating i to the set of reals. Just as we feel more comfortable when we are able to define irrational numbers in terms of rational numbers, so we must define numbers of the form $a + ib$ in terms of the real numbers and justify our somewhat suspect manipulations. First we make the formal definition; then we deduce that our manipulations are in order.

It is easy to do this once you realize that $a + ib$ is completely specified by the ordered pair (a, b) of real numbers (putting aside for the moment the meaning of the symbol i) and so the set $\{a + ib : a, b \in \mathbf{R}\}$ can be put into one-one correspondence with $\mathbf{R} \times \mathbf{R}$. To turn this correspondence into an isomorphism, we need to define suitable addition and multiplication operations on $\mathbf{R} \times \mathbf{R}$. Keeping an eye on the operations we have already found useful, we make the following definition.

Definition

The **complex number system**, denoted by **C**, is the set of all ordered pairs of real numbers (that is, $\mathbf{R} \times \mathbf{R}$) with the operations of **addition** (denoted by $+$) and **multiplication** (denoted by $\times$) which satisfy

$$(a, b) + (c, d) = (a + c, b + d)$$

$$(a, b) \times (c, d) = (ac - bd, bc + ad)$$

for all $(a, b), (c, d) \in \mathbf{C}$.

It is straightforward but tedious to check that $+$ and $\times$ defined as above are both commutative and associative and that $\times$ is distributive over $+$.

These definitions make perfect sense for any real numbers a, b, c and d and present no logical problems about square roots of -1 (no such square roots being mentioned in the definition). Thus we have turned round the earlier process and have defined a complex number as an ordered pair of real numbers which obeys suitable operations $+$ and $\times$.

The notion of equality of complex numbers, comes from equality in $\mathbf{R} \times \mathbf{R}$.

> Two complex numbers, (a, b) and (c, d), are **equal** if and only if $a = c$ and $b = d$.

Having made a proper definition of complex numbers, we now go back and *restore* our original notation, which, as we shall see, is far more convenient than the ordered pair notation. To do this, we note first of all that the subset $\{(a, 0): a \in \mathbf{R}\}$ of $\mathbf{C}$ is isomorphic* to $\mathbf{R}$, with the complex number $(a, 0)$ corresponding to the real number a. It is obviously more convenient to write a instead of $(a, 0)$ and furthermore, because

$$(a, 0) \times (c, d) = (ac, ad),$$

we can write

$$a \times (c, d) = (ac, ad)$$

or, more concisely,

$$a(c, d) = (ac, ad).$$

Then any complex number, (a, b), can be written as

$$(a, b) = (a, 0) + (0, b) = a(1, 0) + b(0, 1).$$

We have already identified $(1, 0)$ with the real number 1; all that remains is to replace $(0, 1)$ by a symbol, i, of course. Thus the symbols

$$(a, b), \quad a(1, 0) + b(0, 1) \quad \text{and} \quad a + bi$$

represent the same complex number, and we write

$$(a, b) = a(1, 0) + b(0, 1) = a + bi.$$

Definition

The symbol i represents $(0, 1) \in \mathbf{C}$.

You may have noticed that we have just written $a + bi$ instead of $a + ib$. It does not matter which you use.

As we remarked earlier, addition and multiplication of complex numbers (regarded as elements of $\mathbf{R} \times \mathbf{R}$) are commutative and associative, and multiplication is distributive over addition. Numbers of the form $a + ib$ can therefore be manipulated according to the ordinary laws of algebra to give

$$(a + ib) + (c + id) = (a + c) + i(b + d)$$

and

$$(a + ib) \times (c + id) = ac + i(bc + ad) + i^2bd.$$

We now have to interpret the symbol i^2. It corresponds to $(0, 1) \times (0, 1)$ and direct application of the definition of multiplication shows that

$$(0, 1) \times (0, 1) = (-1, 0)$$

which we are now abbreviating to -1. We are thus able to write

$$(a + ib) \times (c + id) = (ac - bd) + i(bc + ad),$$

which corresponds to

$$(a, b) \times (c, d) = (ac - bd, bc + ad).$$

---

* "A is isomorphic to B" means that there is an isomorphism from A onto B. See *Unit M100 3, Morphisms*.

Let us summarize these manoeuvres. Practical considerations suggest that a "number", i, with the property $i^2 = -1$ might be very useful. But symbols such as $\sqrt{-1}$ obviously present logical difficulties. Nevertheless, brushing aside these difficulties, we can predict how numbers $a + bi$ would behave. We use this information to define a perfectly logical structure on $\mathbf{R} \times \mathbf{R}$ that is isomorphic to the system we are trying to construct. By taking this structure on $\mathbf{R} \times \mathbf{R}$ as a starting point we are able to identify i with the ordered pair $(0, 1)$ and $a + bi$ with the ordered pair (a, b). The notation $a + bi$ is thus nothing more than an alternative form for (a, b): its advantage is that expressions like $(a + bi) \times (c + di)$ can be manipulated according to the ordinary laws of algebra with i^2 replaced by -1 whenever it appears.

It is often necessary to refer to the first and second components of a complex number and so they are given names.

Definition

> The real number a is called the **real part** of $a + ib$ and the real number b is called the **imaginary part** of $a + ib$.

Notation

We shall write

$$a = \text{Re}(a + ib) \quad \text{and} \quad b = \text{Im}(a + ib).$$

Conventions

A complex number which has imaginary part zero is often called "real", because even though it is not a real number it is, as we have seen, so closely identified with a real number that to insist on the distinction would be pedantic.

We shall find it useful to denote complex numbers by single letters. In particular, just as in real analysis x is often used to denote a variable real number, so, in complex analysis, z is a popular choice of symbol to denote a variable complex number. Furthermore, we shall nearly always implicitly assume that x and y are real and that x, y and z are related by the equations:

$$\text{Re}\, z = x \quad \text{and} \quad \text{Im}\, z = y.$$

In other words,

$$z = (x, y) \quad \text{or} \quad z = x + iy.$$

It is interesting to notice that since two complex numbers are equal if and only if their real parts are equal and their imaginary parts are equal, a single equation connecting complex numbers leads to *two* equations connecting real numbers. Thus if w and z are complex numbers, we can deduce from the single equation

$$w = z$$

that

$$\text{Re}\, w = \text{Re}\, z \quad \text{and} \quad \text{Im}\, w = \text{Im}\, z.$$

This process is usually referred to as *equating real and imaginary parts.*

As you must have realized, despite its name, the imaginary part of a complex number is a perfectly ordinary real number. The name is an unfortunate one because it often leads to the *mistaken* impression that the imaginary part of $a + ib$ is ib rather than b. The term "imaginary" arises from the original concern about complex numbers associated with the symbol $\sqrt{-1}$. The way we have defined it, i is no more imaginary than the number π.

Now that we all know what a complex number is, we need to agree on a list of properties of $\mathbf{C}$ that we can take as a starting point for complex analysis. We already know that $\mathbf{C}$ has some of the properties of $\mathbf{R}$ (such as associativity and commutativity of addition and of multiplication, and distributivity of multiplication over addition). In fact, all the familiar arithmetic (field)* properties of **R** carry over to **C** as theorems, in the sense that they can be proved from the properties of **R**. In the list that follows z, z_1, z_2, z_3 are any elements of **C**: in other words the properties hold for all $z, z_1, z_2, z_3 \in \mathbf{C}$.

Properties of C

(i) Associative law for addition:

$$z_1 + (z_2 + z_3) = (z_1 + z_2) + z_3.$$

(ii) Existence of an additive identity (denoted by 0):

there is $0 \in \mathbf{C}$ such that $z + 0 = z$.

(iii) Existence of additive inverses:

there is a $(-z) \in \mathbf{C}$ such that $z + (-z) = 0$.

(iv) Commutative law for addition:

$$z_1 + z_2 = z_2 + z_1.$$

(v) Associative law for multiplication:

$$z_1 \times (z_2 \times z_3) = (z_1 \times z_2) \times z_3.$$

(vi) Existence of a multiplicative identity (denoted by 1):

there is $1 \in \mathbf{C}$ such that $z \times 1 = z$.

(vii) Existence of multiplicative inverses of non-zero numbers:

if $z \neq 0$, there is a $z^{-1} \in \mathbf{C}$ such that $z \times z^{-1} = 1$.

(viii) Commutative law for multiplication:

$$z_1 \times z_2 = z_2 \times z_1.$$

(ix) Distributive law:

$$z_1(z_2 + z_3) = z_1 z_2 + z_1 z_3.$$

(Note that in (ix) we have suppressed the multiplication symbol. This is common practice, as is the use of $\cdot$ for $\times$.)

We have already mentioned properties (i), (iv), (v), (viii) and (ix). Properties (ii), (iii), (vi) and (vii) are easily checked.

The additive identity, 0, is $(0, 0) = 0 + 0i$ and the additive inverse of $z = x + iy$ is $-x + i(-y)$ which is written $-x - iy$ or $-z$. As with the real numbers, we define the **difference of two complex numbers** by

$$\begin{aligned}(a + ib) - (c + id) &= (a + ib) + (-c - id)\\ &= (a - c) + i(b - d)\end{aligned}$$

or

$$w - z = w + (-z).$$

The multiplicative identity, 1, is $(1, 0) = 1 + 0i$, and so if $w = u + iv$ is the multiplicative inverse of $z = x + iy$ then

$$zw = 1,$$

that is,

$$(x + iy)(u + iv) = 1$$

and so

$$xu - yv + i(yu + xv) = 1.$$

---

* The concept of a field is discussed fully in M. Spivak, *Calculus*. A field is a set with two closed binary operations (+ and ×) for which properties (i)–(ix) are satisfied.

Thus, equating real and imaginary parts, we have

$$xu - yv = 1 \quad \text{and} \quad yu + xv = 0,$$

giving

$$u = \frac{x}{x^2 + y^2} \quad \text{and} \quad v = \frac{-y}{x^2 + y^2},$$

since x and y cannot both be zero (because $z \neq 0$). Thus the multiplicative inverse of $x + iy$ is

$$\frac{x}{x^2 + y^2} - \frac{iy}{x^2 + y^2}.$$

All the other useful facts about **C**, such as the uniqueness of 0, $wz = 0$ if and only if $w = 0$ or $z = 0$, and so on, follow from the nine properties as for real numbers. The importance of this is that all the usual identities, such as the difference of two squares, the sum of a (finite) geometric series, the binomial theorem for a positive integer index go through as in real analysis. But now we come to the differences between real analysis and complex analysis. Most real analysis courses rely heavily on other properties of the real numbers—properties that tell us that the real numbers are *ordered* and enable us to define inequalities on the real numbers. (See **Spivak***, pages 9 and 10.) This is a rather special use of the term "ordered". We talk of a field being *ordered* if we can identify a subset of "positive elements" such that one and only one of the following holds for each element, a, of the field: a is positive; $-a$ is positive; a is zero (the trichotomy law); *and*, if a and b are positive then so are $a + b$ and $a \times b$. (The set of real numbers is an example of an ordered field.) In this sense of the word "ordered", clearly there are problems with trying to turn **C** into an ordered field. For how could we define "positive" and "negative" subsets in **C** which have the properties we would expect?

The concept of an ordered field is an attempt to characterize the special ordering of the real numbers and it is fairly clear that the two-dimensional nature of the complex numbers is going to rule out the possibility of their being ordered in this sense. But suspicion is not enough; we should *prove* that **C** is not an ordered field.

It is easy enough to show that in any structure satisfying (i)–(ix) and in which we *can* talk of "positive" and "negative" elements, the square of any non-zero element is positive. This is precisely the property we set out not to have in constructing the complex numbers! (For, if a is positive then $a \times a = a^2$ must be positive, by the trichotomy law, and if a is negative then $-a$ is positive and so $(-a) \times (-a) = a^2$ is positive, again by the trichotomy law.) If we wanted in some way to extend the concepts of positive and negative from **R** to **C** we would presumably require $(1, 0)$ to be positive, in which case, by the trichotomy law, $(-1, 0)$ would be negative. Then $(0, 1)$ is non-zero, but $(0, 1)^2 = (-1, 0)$ is negative (and so we have the square of a non-zero element being negative). If we decided to get round this difficulty by calling $(-1, 0)$ "positive", it follows that $-(-1, 0) = (1, 0)$ must be non-zero; also, then $(1, 0)^2 = (1, 0)$ is negative (and again we have the square of a non-zero element being negative). We have shown that it is not possible to find two classes which we could justifiably call "positive" and "negative". It was precisely the existence of these two classes in **R** that enabled us to define inequalities in **R**.

We therefore come to the conclusion that the relation $<$ does not carry over from the real numbers to the complex numbers.

You may remember that it is precisely the relation $<$ that enables us to talk of least upper bounds (see **Spivak**)—and so in moving from real analysis to complex analysis we have to throw overboard the least upper bound property that is so useful in real analysis.

Fortunately things are not as black as they might seem, because the proofs in which least upper bounds are used most frequently in real analysis can also be done by the use of the technique by repeated bisection. This relies on a theorem (proved from the least upper bound property) called the Nested Intervals Theorem (see **Spivak**). Happily, this theorem has an analogue in complex analysis called the Nested Rectangles Theorem. We shall prove this theorem in Section 1.5 and it will occupy a fundamental role in this course similar to that occupied by the least upper bound property in most real analysis courses.

---

* This form of reference is used for M. Spivak, *Calculus*. See conventions on page 6.

Summary

In this section we have seen that the complex number system is the set $\mathbf{R} \times \mathbf{R}$ together with the operations $+$ and $\times$ defined by

$$(a, b) + (c, d) = (a + c, b + d)$$

$$(a, b) \times (c, d) = (ac - bd, ad + bc).$$

From this, one can justify the performance of ordinary algebraic operations on expressions of the form $a + ib$ where $a, b \in \mathbf{R}$ and i is an undefined symbol with the property that i^2 is replaced by -1.

In the complex number system, any quadratic expression can be factorized into two linear factors.

The real part of a complex number $z = x + iy$ is x and we write

$$\operatorname{Re} z = \operatorname{Re}(x + iy) = x.$$

The imaginary part of a complex number $z = x + iy$ is y and we write

$$\operatorname{Im} z = \operatorname{Im}(x + iy) = y.$$

Two complex numbers are equal if and only if their real parts are equal and their imaginary parts are equal.

The complex number system obeys the usual algebraic laws of manipulation of numbers but the inequality relation, $<$, does not carry over from the real numbers to the complex numbers.

Examples

1. $(3 + i)(4 - i) = 12 - i^2 + 4i - 3i = 13 + i.$
2. $(6 + 7i) - (3 - i) = 3 + 8i.$
3. $(x + iy)(x - iy) = x^2 + y^2.$
4. $\operatorname{Re}(6 - 4i) = 6,$ and $\operatorname{Im}(6 - 4i) = -4.$
5. If

   $$xy - (x^2 + y^2)i = 1 - 3i \quad (x, y \in \mathbf{R}),$$

   then, by equating real and imaginary parts, we have

   $$xy = 1 \quad \text{and} \quad x^2 + y^2 = 3.$$

   Thus the point (x, y) lies on the hyperbola $xy = 1$ and on the circle $x^2 + y^2 = 3$ and can take four possible values, corresponding to the points of intersection of the hyperbola and the circle.

Self-Assessment Questions

Self-assessment questions are intended to test your immediate comprehension of a reading section. If you have difficulty with them you should read again the appropriate parts of the material. Before checking the solution to any part of a question, you should work all the parts of that question. The solutions follow the questions.

1. In each of the following cases, indicate whether the statement is True or False.

   (i) $\text{Re}(3 + 2i) = 3$.

   (ii) $\text{Im}(3 + 2i) = 2i$.

   (iii) If $z_1 z_2$ is a real number, then z_1 and z_2 must be real.

   (iv) If $z_1 + z_2$ is a real number, then z_1 and z_2 must be real.

2. Express each of the following in the form $a + bi$ where a and b are real numbers.

   (i) $(3 + 2i) + (2 - i)$.

   (ii) $(4 + i) - (2 - i)$.

   (iii) $(3 + 2i)(2 - i)$.

   (iv) $\text{Re}(6 - 7i)$.

   (v) $\text{Im}(3 - 2i)$.

3. Find x and y if $(x + 3i)(2 + yi) = 5 + 5i$, and $x, y \in \mathbf{R}$.

Solutions

1. (i) True.

   (ii) False, $\text{Im}(3 + 2i) = 2$.

   (iii) False, consider $z_1 = 1 + i$, $z_2 = 1 - i$.

   (iv) False, consider $z_1 = x + iy$ and $z_2 = x - iy$.

2. (i) $5 + i$. (ii) $2 + 2i$. (iii) $8 + i$. (iv) $6 + 0i = 6$.
   (v) $-2 + 0i = -2$ (not $2i$).

3. We have

$$(2x - 3y) + (6 + xy)i = 5 + 5i.$$

Equating real and imaginary parts,

$$2x - 3y = 5 \quad \text{and} \quad 6 + xy = 5.$$

Thus, $x = 1$ and $y = -1$ or $x = \frac{3}{2}$ and $y = -\frac{2}{3}$.

1.2 PROBLEMS

The solutions immediately follow the last problem.

1. Simplify each of the following expressions into the form $a + ib$, where a and b are real.

   (i) $(1 + 6i)(1 - 3i)$.

   (ii) $(1 + i)^3$.

   (iii) $(1 - i)(1 + i)$.

   (iv) $(3 + 2i)(3 - 2i)$.

   (v) $(3 + 7i) - (2 + i)$.

2. Find all possible values of z satisfying

   (i) $z^2 = -9$.

   (ii) $z^2 = -k^2$, k real.

3. Solve the equation

$$(1 + i)(x + iy) = -2 + i$$

where $x, y \in \mathbf{R}$, by deducing simultaneous equations in x and y.

As a consequence of the carry over from **R** to **C** of the usual arithmetic properties, many of the familiar definitions and results of the algebra of real numbers extend easily to complex numbers. For example, an expression

$$p(z) = a_0 + a_1 z + \cdots + a_n z^n$$

where z and the a_i's are complex is called a **polynomial** (in z). If $a_n \neq 0$, n is called the *degree* of the polynomial.

---

4. (i) Use mathematical induction to prove that if n is any positive integer, there are positive real numbers $c_1, \ldots, c_n$ (each dependent upon n) such that

$$(1 + z)^n = 1 + c_1 z + c_2 z^2 + \cdots + c_n z^n$$

   for all complex numbers z.

   Identify the coefficients.

   (ii) Generalize the results of (i) to $(w + z)^n$.

5. (i) If n is a positive integer, show that

$$\frac{1 - z^n}{1 - z} = 1 + z + z^2 + \cdots + z^{n-1}, \quad (z \neq 1).$$

   (Hint: Denote the right-hand side by $p(z)$, and consider $p(z) - zp(z)$.)

   (ii) Deduce that

$$\frac{z^n - w^n}{z - w} = z^{n-1} + z^{n-2}w + \cdots + w^{n-1}, \quad (z \neq w).$$

   (Hint: See the solution to part (ii) of Problem 4).

6. (i) A polynomial $p(z)$ is *divisible* by $z - z_0$ if a polynomial $q(z)$ exists such that $p(z) = (z - z_0)q(z)$. Show that for any positive integer n, $z^n - z_0^n$ is divisible by $z - z_0$.

   (ii) If $p(z)$ is a polynomial, show that the polynomial $p(z) - p(z_0)$ is divisible by $z - z_0$.

   (iii) Show that

$$p(z) \text{ is divisible by } z - z_0 \Leftrightarrow p(z_0) = 0.$$

(iv) The Fundamental Theorem of Algebra states that if $p(z)$ is a non-constant polynomial, then the equation $p(z) = 0$ has a (complex) root. We shall prove this theorem later in the course. Assuming the theorem, prove that every polynomial of degree $n > 0$ can be factorized into linear factors in the form

$$p(z) = k(z - z_0)(z - z_1)(z - z_2)\cdots(z - z_{n-1}).$$

When we wish to be explicit about the operations defined on **C** which are under consideration we sometimes write $(\mathbf{C}, +, \times)$, $(\mathbf{C}, +)$ or $(\mathbf{C}, \times)$.

---

7. Show that there is an isomorphism from $(\mathbf{C}, +, \times)$ to the set of matrices of the form $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$, $a, b \in \mathbf{R}$, under the operations of matrix addition and multiplication.

8. (i) Show that there is an isomorphism from $(\mathbf{C}, +)$ to the set of two-dimensional geometric vectors with the operation of vector addition.

   (ii) Show that the mapping $z \longrightarrow z_0 + z$ from **C** to itself, where z_0 is a fixed complex number, is equivalent to a translation of $\mathbf{R} \times \mathbf{R}$. (A mapping ϕ for which $\phi(x, y) = (x + \alpha, y + \beta)$, where α and β are fixed, is a translation of $\mathbf{R} \times \mathbf{R}$.)

9. Show that if $z_0 = (x_0, y_0)$ is a fixed complex number, the mapping from **C** to itself defined by $f: z \longrightarrow z_0 z$ is linear; that is, that

$$f(z_1 + z_2) = f(z_1) + f(z_2)$$
$$f(cz) = cf(z)$$

for all $z_1, z_2, c, z \in \mathbf{C}$.

Solutions

1. (i) $19 + 3i$.

   (ii) $$\begin{aligned}(1 + i)^3 &= (1 + i)(1 + i)(1 + i)\\ &= (1 + i)(0 + 2i)\\ &= -2 + 2i.\end{aligned}$$

   (iii) 2. (iv) 13. (v) $1 + 6i$.

2. (i) If $z = x + iy$, the equation $z^2 = -9$ becomes

$$(x^2 - y^2) + 2xyi = -9.$$

   Thus $x^2 - y^2 = -9$ and $2xy = 0$. From $xy = 0$ we deduce that x or y is zero. From $x^2 - y^2 = -9$ we deduce that y cannot be zero. Thus $x = 0$, $y = \pm 3$ and so $z = \pm 3i$.

   (ii) $z = \pm ki$.

3. The equation is equivalent to $x - y + i(x + y) = -2 + i$. Thus $x - y = -2$ and $x + y = 1$, whence $x = -\frac{1}{2}$, $y = \frac{3}{2}$.

4. (i) For $n = 1$, $(1 + z)^1 = 1 + z$ and $c_1 = 1$.
   If the result is true for $n = k$,

$$\begin{aligned}(1 + z)^{k+1} &= (1 + z)^k(1 + z)\\ &= (1 + c_1 z + \cdots + c_k z^k)(1 + z)\end{aligned}$$

   where $c_1, c_2, \ldots, c_k$ are positive real numbers (dependent on k). (In particular, note that the c_1 here is not the c_1 in the first line of this solution. The present c_1 is associated with $n = k$; the first one is associated with $n = 1$.) We thus deduce

$$\begin{aligned}(1 + z)^{k+1} = 1 &+ (c_1 + 1)z + (c_2 + c_1)z^2 + \cdots\\ &+ (c_k + c_{k-1})z^k + (c_k)z^{k+1},\end{aligned}$$

   and so the coefficients are real and positive.

Furthermore, they satisfy the familiar property of the binomial coefficients, often expressed as Pascal's Triangle, and so

$$(1+z)^n = 1 + nz + \frac{n(n-1)}{2}z^2 + \cdots + nz^{n-1} + z^n.$$

(ii) Similarly,

$$(w+z)^n = w^n + nw^{n-1}z + \frac{n(n-1)}{2}w^{n-2}z^2 + \cdots + nwz^{n-1} + z^n.$$

(You can derive this from the previous result by replacing z by z/w, $w \neq 0$, and then multiplying by w^n. The case $w = 0$ is dealt with independently, the result being trivial.)

5. (i) Since $p(z) - zp(z) = 1 - z^n$, we have

$$p(z)(1-z) = 1 - z^n,$$

from which,

$$\frac{1-z^n}{1-z} = p(z), \text{ since } z \neq 1.$$

(ii) First, note that $z \neq w$, that is $z/w \neq 1$. Then replacing z by z/w, $w \neq 0$, in (i) gives

$$\frac{1-\left(\frac{z}{w}\right)^n}{1-\frac{z}{w}} = 1 + \left(\frac{z}{w}\right) + \left(\frac{z}{w}\right)^2 + \cdots + \left(\frac{z}{w}\right)^{n-1},$$

provided $z/w \neq 1$, that is $z \neq w$. Thus

$$\left(\frac{w^n - z^n}{w-z}\right)\frac{1}{w^{n-1}} = 1 + \left(\frac{z}{w}\right) + \left(\frac{z}{w}\right)^2 + \cdots + \left(\frac{z}{w}\right)^{n-1}.$$

The result follows from multiplication by w^{n-1}, and the fact that $(w^n - z^n)/(w-z) = (z^n - w^n)/(z-w)$. The case $w = 0$ is dealt with independently: again the result is trivial.

6. (i) From Problem 5(ii),

$$q(z) = z^{n-1} + z^{n-2}z_0 + z^{n-3}z_0 + \cdots + z_0^{n-1};$$

so $q(z)$ exists and $z^n - z_0^n$ is divisible by $z - z_0$.

(ii) If $p(z) = a_0 + a_1z + \cdots + a_nz^n$,

$$p(z) - p(z_0) = a_1(z - z_0) + a_2(z^2 - z_0^2) + \cdots + a_n(z^n - z_0^n).$$

From (i), each of the terms on the right can be written in the form $a_i(z^i - z_0^i) = a_i(z - z_0)q_i(z)$ where $q_i(z)$ is a polynomial. Thus

$$p(z) - p(z_0) = (z - z_0)q(z)$$

where

$$q(z) = \sum_{i=1}^{n} a_iq_i(z).$$

(iii) If $p(z)$ is divisible by $z - z_0$, there is a polynomial $q(z)$ such that $p(z) = (z - z_0)q(z)$ and so $p(z_0) = 0$.
If $p(z_0) = 0$, the equation $p(z) - p(z_0) = (z - z_0)q(z)$ becomes $p(z) = (z - z_0)q(z)$ and so $p(z)$ is divisible by $z - z_0$.

(iv) From the Fundamental Theorem of Algebra, there is a number z_n for which $p(z_n) = 0$ and so parts (iii) and (i) enable us to write

$$p(z) = (z - z_n)q(z).$$

The polynomial $q(z)$ is of degree $n - 1$ and the argument can now be applied to $q(z)$. The result follows by induction.

7. The mapping $f: a + ib \longrightarrow \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ is clearly one-one and onto. Also

$$\begin{aligned} f(a + ib) + f(c + id) &= \begin{bmatrix} a & b \\ -b & a \end{bmatrix} + \begin{bmatrix} c & d \\ -d & c \end{bmatrix} \\ &= \begin{bmatrix} a + c & b + d \\ -(b + d) & a + c \end{bmatrix} \\ &= f\big(a + c + i(b + d)\big) \\ &= f\big((a + ib) + (c + id)\big); \end{aligned}$$

and

$$\begin{aligned} f(a + ib) \times f(c + id) &= \begin{bmatrix} a & b \\ -b & a \end{bmatrix}\begin{bmatrix} c & d \\ -d & c \end{bmatrix} \\ &= \begin{bmatrix} ac - bd & ad + bc \\ -(ad + bc) & ac - bd \end{bmatrix} \\ &= f\big((ac - bd) + i(ad + bc)\big) \\ &= f\big((a + ib) \times (c + id)\big). \end{aligned}$$

Thus f is an isomorphism.

(As you would expect, the inverse matrix corresponds to the multiplicative inverse of $a + ib$.)

8. (i) The mapping $f: a + ib \longrightarrow a\underset{\rightarrow}{i} + b\underset{\rightarrow}{j}$ is clearly one-one and onto from **C** to the set of two-dimensional geometric vectors and

$$\begin{aligned} f\big((a + ib) + (c + id)\big) &= f\big((a + c) + i(b + d)\big) \\ &= (a + c)\underset{\rightarrow}{i} + (b + d)\underset{\rightarrow}{j} \\ &= (a\underset{\rightarrow}{i} + b\underset{\rightarrow}{j}) + (c\underset{\rightarrow}{i} + d\underset{\rightarrow}{j}) \\ &= f(a + ib) + f(c + id). \end{aligned}$$

Note that we have used the Foundation Course notation $\underset{\rightarrow}{i}$ and $\underset{\rightarrow}{j}$ for geometric vectors of unit length in the direction of the x-axis and y-axis respectively.

(This isomorphism gives a useful visualization of addition and subtraction of complex numbers and also suggests how they may usefully be represented geometrically. This is taken up in the next section.)

(ii) If $z_0 = x_0 + iy_0$, the effect of the mapping is to map (x, y) to $(x_0 + x, y_0 + y)$, and it is therefore a translation.

9. We have

$$\begin{aligned} f(z_1 + z_2) = z_0(z_1 + z_2) &= z_0 z_1 + z_0 z_2 \\ &= f(z_1) + f(z_2); \end{aligned}$$

$$\begin{aligned} f(cz) = z_0(cz) &= c(z_0 z) \\ &= cf(z). \end{aligned}$$

Thus f is linear.

1.3 THE COMPLEX PLANE

In this section we shall develop the correspondence between **C** and **R** × **R** by obtaining a geometric representation of elements of **C** and operations on **C**. We shall define the polar form of a complex number and the modulus and argument of a complex number. We shall see that not only does $\sqrt{-1}$ have some meaning in **C** but that every non-zero complex number has exactly n nth roots in **C**.

* * * * * * * *

We have seen in Section 1.1 that the complex number system is obtained by defining arithmetic operations on the set **R** × **R**. We also know that elements of **R** × **R** can be represented as points in a plane. It seems reasonable to ask what insight can be obtained by representing complex numbers as points of the plane.

One obvious way of making the representation is to use Cartesian coordinates. Then the number $x + iy$ is represented by the point with coordinates (x, y). Thus if a point P represents a complex number z, then its x coordinate is the real part of z and its y coordinate is the imaginary part of z (hence the choice of notation $z = x + iy$ made in Section 1.1).

The horizontal axis is called the *real axis* and the vertical axis is called the *imaginary axis.*

This way of representing complex numbers by points in the plane is used so often that we shall use phrases such as "the point $x + iy$" and the "the point z" as convenient shorthands for "the point representing z" and we shall speak of the **complex plane.** Fig. 1 shows several points in the complex plane

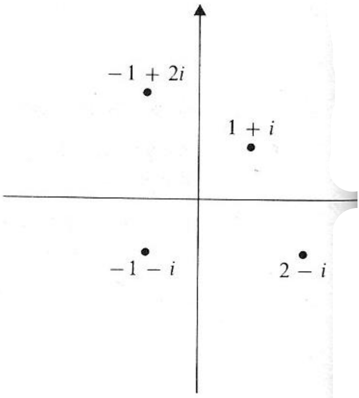


Fig. 1

As we saw in Problem 8 of Section 1.2 (and it is fairly obvious anyway), the geometric representation can be taken a stage further. Complex numbers also add in the same way as number pairs regarded as vectors and so they add according to the parallelogram law (Fig. 2).

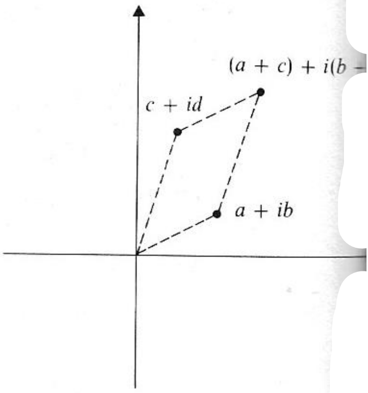


Fig. 2

Now points in the plane can also be represented by polar coordinates and this suggests another way of specifying a complex number.

If you have forgotten about polar coordinates, a short revision will not come amiss. When we use polar coordinates—instead of specifying a point in a plane by the distance from two lines, as in Cartesian coordinates—we specify a point P in the plane by its distance (r) from a fixed point O (the origin) and the angle (θ) that OP makes with a fixed half-line (the initial line) which has one end at O (Fig. 3). (At present we consider only points P different from O.)

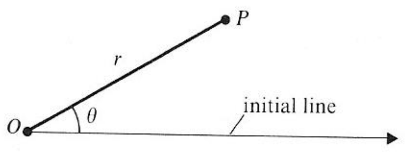


Fig. 3

As is usual in mathematics, the angle θ is measured in an anticlockwise direction from the initial line (Fig. 4).

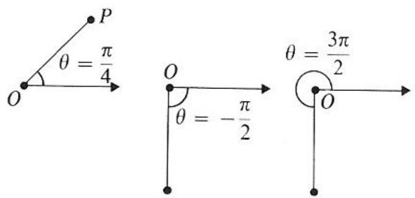


Fig. 4

One snag that does arise with polar coordinates is that they do *not* give a one-one correspondence with the points of the plane. Thus a point with polar coordinates (r, θ) also has polar coordinates $(r, \theta + 2k\pi)$ where k is any integer. But polar coordinates (r, θ) do specify a *unique* complex number, and from the relationship between polar coordinates and Cartesian coordinates we know that this number is $r \cos \theta + ir \sin \theta$ (Fig. 5). So whether we use polar coordinates or Cartesian coordinates, a point in the plane specifies a unique complex number.

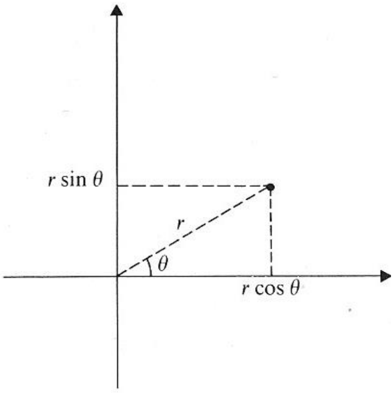


Fig. 5

It is often useful to represent a complex number in terms of the polar coordinates of the corresponding point as in the following definition.

Definition

A **polar form** of the non-zero complex number $z = x + iy$ is $r(\cos \theta + i \sin \theta)$ where $r = \sqrt{x^2 + y^2}$, $\sin \theta = y/r$ and $\cos \theta = x/r$. The positive number r is called the **modulus** of z and θ is **an argument** of z. We denote the modulus of z by $|z|$, and so $r = |z| = |x + iy|$. For the special case of $z = 0$, we define $|z| = 0$, but we do not need to define what is meant by an argument of 0.

As we know, there will be some difficulty in specifying arguments. This is manifested by the fact that the relations $\sin \theta = y/r$, $\cos \theta = x/r$ specify θ only to within an integer multiple of 2π. (Notice that the equation $\tan \theta = y/x$ is even less specific.)

Given any non-zero complex number, z, an argument can be any element from a whole set of numbers. For example, an argument of $1 + i$ can be any element from $\{\pi/4, 9\pi/4, -7\pi/4, 17\pi/4, -15\pi/4, \ldots\}$. Nevertheless, any one element from this set is sufficient to identify the set and whatever the number $z(\neq 0)$ there is only one of its arguments in the interval $(-\pi, \pi]$.

Definition

If θ is an argument of z and $\theta \in (-\pi, \pi]$, θ is the **principal value of the argument** of z, and we write $\theta = \operatorname{Arg} z$. (Note the capital A.)

Thus, for example, $\operatorname{Arg}(1 + i) = \pi/4$ (Fig. 6).

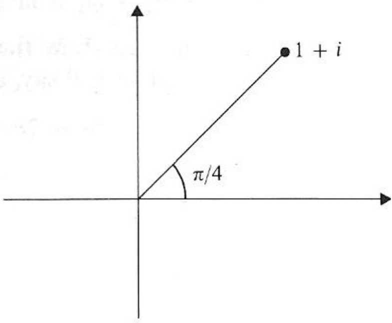


Fig. 6

It would have been just as reasonable to have chosen the interval $[0, 2\pi)$* or $(0, 2\pi]$ or $[-\pi, \pi)$ for $\operatorname{Arg} z$, but we must settle for one option—we happen to have chosen $(-\pi\ \pi]$: this choice is not ideal for all purposes, but our reasons will be discussed in *Unit 3, Differentiation.*

We shall see many times in the course how convenient the polar form of a complex number can be and so we investigate now the effect of the arithmetic operations on the polar forms of complex numbers.

Addition and subtraction of complex numbers given in polar form is not a pleasant affair if the answers are also required to be in polar form. If

$$z = r \cos \theta + ir \sin \theta$$

and

$$w = \rho \cos \phi + i\rho \sin \phi,$$

then

$$z + w = r \cos \theta + \rho \cos \phi + i(r \sin \theta + \rho \sin \phi)$$

and that is about as far as you can get unless you are a wizard at trigonometry. (The polar form would be $p \cos \psi + ip \sin \psi$ where p and ψ are expressions in r, ρ, θ and ϕ.) Of course, in any particular case you can work out the polar form of the answer, but the general result is not worth working out, let alone remembering.

---

* In the Foundation Course we chose $[0, 2\pi)$.

On the other hand, multiplication is easy because

$$\begin{aligned} zw &= r\rho[\cos\theta\cos\phi - \sin\theta\sin\phi + i(\sin\theta\cos\phi + \cos\theta\sin\phi)] \\ &= r\rho\cos(\theta + \phi) + ir\rho\sin(\theta + \phi). \end{aligned}$$

Thus

$$|zw| = r\rho = |z| \cdot |w|,$$

and *an* argument of zw can be obtained just by adding arguments of z and w. If we want to work with principal values, there is one minor difficulty: even if θ and ϕ are in $(-\pi, \pi]$, when we add θ and ϕ we might well get an answer outside $(-\pi, \pi]$. So if we want to work with principal values we shall have to make sure to get the answer back into $(-\pi, \pi]$. Thus although $\operatorname{Arg} z + \operatorname{Arg} w$ is an argument of zw we *may* have to add or subtract 2π to obtain $\operatorname{Arg}(zw)$.

One possible snag that may have occurred to you is this: if θ and ϕ are arguments of z and w respectively, is $\theta + \phi$ an argument of zw *whatever* θ and ϕ we choose? Well, the relation where θ_1 is related to θ_2 if $\theta_1 = \theta_2 + 2k\pi$, k an integer, is an equivalence relation on $\mathbf{R}$ and all the arguments of a complex number will constitute an equivalence class. (See *Unit M100 19, Relations.*) We are asking whether we can combine equivalence classes (add arguments), in other words whether the equivalence relation is compatible with addition. Can we write

$$\{\text{arguments of } z\} + \{\text{arguments of } w\} = \{\text{arguments of } zw\}?$$

We can, of course, because if θ_1 and θ_2 are in the same equivalence class and ϕ_1 and ϕ_2 are in the same equivalence class, there are integers k and m such that

$$\theta_1 = \theta_2 + 2k\pi \quad \text{and} \quad \phi_1 = \phi_2 + 2m\pi,$$

and so

$$\theta_1 + \phi_1 = \theta_2 + \phi_2 + 2(k + m)\pi,$$

that is, $\theta_1 + \phi_1$ is in the same equivalence class as $\theta_2 + \phi_2$.

We can also show that if α is a fixed argument of zw then we can choose an argument of z, θ say, and an argument of w, ϕ say, such that $\theta + \phi = \alpha$.

Let $\alpha = \operatorname{Arg} zw + 2n\pi$, where n is a fixed integer. Certainly,

$$\operatorname{Arg} zw = \operatorname{Arg} z + \operatorname{Arg} w + 2k\pi \quad \text{for } k = -1, 0 \text{ or } 1;$$

and so

$$\alpha = \operatorname{Arg} z + \operatorname{Arg} w + 2(k + n)\pi.$$

So choose

$$\theta = \operatorname{Arg} z + 2p\pi, \quad \phi = \operatorname{Arg} w + 2q\pi$$

where $p + q = k + n$.

We define **division of complex numbers** (as for real numbers) as the inverse of multiplication, and so it is no surprise that polar forms are also very suitable for division. If z^{-1} is the number such that

$$zz^{-1} = 1 \qquad (z \neq 0)$$

then we define $w \div z$ by

$$w \div z = w \times z^{-1}.$$

As for real numbers, we write

$$z^{-1} = \frac{1}{z} \quad \text{and} \quad w \div z = \frac{w}{z}.$$

We now ask how to calculate $1/z$ when $z\ (\neq 0)$ is given. If θ is an argument of z and ϕ an argument of $1/z$, then $\theta + \phi$ is an argument of $z \times \frac{1}{z} = 1$. Since 0 is an argument of 1, we must be able to choose ϕ so that

$$\theta + \phi = 0,$$

and so $-\theta$ is an argument of $1/z$. In fact, unless z is a negative real number, we can use principal values without trouble:

$$\operatorname{Arg}\frac{1}{z} = -\operatorname{Arg} z \qquad (z \neq -k,\ k > 0).$$

(Note that a statement like $k > 0$ implies that k is real.)

If $z = -k$, $k > 0$, then $\operatorname{Arg} z = \pi$ and $\operatorname{Arg} 1/z = \pi$.

What about $|1/z|$? We know that $|1| = 1$, and so

$$|z| \cdot \left|\frac{1}{z}\right| = \left|z \times \frac{1}{z}\right| = |1| = 1.$$

Thus

$$\left|\frac{1}{z}\right| = \frac{1}{|z|}.$$

We can now interpret division quite easily in terms of polar forms:

$$\left|\frac{w}{z}\right| = \frac{|w|}{|z|},$$

and if θ and ϕ are arguments of z and w respectively, then $\phi - \theta$ is an argument of w/z. More concisely,

$$\frac{w}{z} = \frac{\rho(\cos\phi + i\sin\phi)}{r(\cos\theta + i\sin\theta)} = \frac{\rho}{r}[\cos(\phi - \theta) + i\sin(\phi - \theta)].$$

To summarize: each of the arithmetic operations on **C** has a geometric interpretation. If $z = x + iy$ is any complex number, corresponding to a point P, and if $w = u + iv = \rho(\cos\phi + i\sin\phi)$ then:

(i) the effect of adding w to z is to *translate* P to the point $(x + u, y + v)$;
the effect of subtracting w from z is to *translate* P to the point $(x - u, y - v)$;

(ii) the effect of multiplying z by w is to *rotate* OP through the angle ϕ and "*stretch*" OP by a factor ρ;
the effect of dividing z by w is to rotate OP through the angle $-\phi$ and "*stretch*" OP by a factor $1/\rho$.

The fact that when multiplying complex numbers their arguments are added suggests an interesting relationship between numbers z and w for which

$$|z| = |w| \quad \text{and} \quad \operatorname{Arg} z = -\operatorname{Arg} w.$$

Multiplying these two numbers gives a number with a zero argument—a real number (Fig. 7). What is more, the modulus of this number is $|z| \cdot |w|$ and this is just $|z|^2$. In polar form we see that if

$$z = r(\cos\theta + i\sin\theta),$$

$$w = r[\cos(-\theta) + i\sin(-\theta)] = r(\cos\theta - i\sin\theta)$$

and so if $z = x + iy$, then $w = x - iy$.

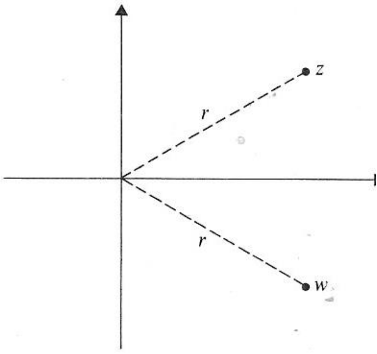


Fig. 7

The relationship between z and w is important and leads us to the following definition.

Definition

If $z = x + iy$, the number $x - iy$ is the **conjugate** of z, and we write $\bar{z} = x - iy$. (Note that many authors write z^* for $\bar{z}$.)

We have just seen that $\bar{z}$ has the important property that $z\bar{z} = x^2 + y^2 = |z|^2$.

This result is useful for dealing with quotients in Cartesian form, for an expression like

$$\frac{a + ib}{c + id}$$

could be simplified without resorting to polar form if we could make the denominator real. We now know how to do that. Since $(c - id)/(c - id) = 1$, we can write

$$\begin{aligned}\frac{a + ib}{c + id} &= \frac{(a + ib)(c - id)}{(c + id)(c - id)}\\ &= \frac{ac + bd + i(bc - ad)}{c^2 + d^2}\\ &= \frac{ac + bd}{c^2 + d^2} + i\frac{bc - ad}{c^2 + d^2}.\end{aligned}$$

The details of this expression are *not* worth remembering, but the technique is: multiply numerator and denominator by the conjugate of the denominator. It is often useful to remember that a denominator z becomes $|z|^2$ after the process.

Perhaps the most impressive payoff from writing complex numbers in polar form arises in the problem of solving an equation of the form

$$z^n = a$$

where a is a given non-zero complex number. Now it is easy to deduce from our knowledge of multiplication of polar forms that

$$|z|^n = |z^n|$$

and that if θ is an argument of z then $n\theta$ is an argument of z^n. (See Problem 7 of Section 1.4.) So if $z^n = a$, z must be a number whose modulus is an nth root of a and which has an argument which is $1/n$ times an argument of a.

Example 1

Find a value of z such that $z^4 = 16i$.

Solution

Since $|16i| = 16$, we have $|z| = 16^{\frac{1}{4}} = 2$.

An argument of $16i$ is $\frac{\pi}{2}$, so an argument of z is $\frac{\pi}{8}$, because $4 \times \frac{\pi}{8} = \frac{\pi}{2}$. So $2\left(\cos\frac{\pi}{8} + i\sin\frac{\pi}{8}\right)$ is a possible value of z. But $4 \times \left(\frac{\pi}{8} + \frac{\pi}{2}\right) = \frac{\pi}{2} + 2\pi$ and this is also an argument of a, so

$$2\left[\cos\left(\frac{5\pi}{8}\right) + i\sin\left(\frac{5\pi}{8}\right)\right]$$

is another possible value of z.

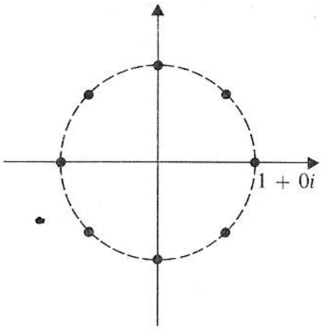
This example clearly suggests that we need a *systematic* method of finding *all* the solutions of the equation $z^n = a$, $a \neq 0$. We can see how to do this by considering the simpler problem of finding all the solutions of $w^n = 1$. Since $|1| = 1$, we must have $|w|^n = 1$ and if ϕ is an argument of w, $n\phi$ must be a multiple of 2π because 0 is an argument of 1. In other words, $|w| = 1$ and any of the values $\frac{k2\pi}{n}$ $(k = 1, 2, \ldots, n)$ can be an argument of w—each leading to a different value of w. (If k takes integer values beyond n, we get repetitions of the values of

$$\cos\frac{k2\pi}{n} + i\sin\frac{k2\pi}{n}$$

that have already been obtained.) Thus in the set of complex numbers, the number 1 has n nth roots:

$$\cos\frac{k2\pi}{n} + i\sin\frac{k2\pi}{n}, \quad k = 1, 2, \ldots, n.$$

The points representing the roots lie equally spaced around the unit circle. Fig. 8 shows the roots for $n = 8$.



Solutions of $w^8 = 1$

Fig. 8

Now let us get back to the original problem: $z^n = a$, $a \neq 0$. We do not seem to have any difficulty in finding one solution of this equation; the problem is how to find the whole set of solutions. Let us suppose we have one solution, z_0, so that $z_0^n = a$. We now use the solution to the simplified problem, $w^n = 1$. Suppose w is a solution of this problem, then $(wz_0)^n = w^n z_0^n = 1a = a$, and so wz_0 is a solution to the original problem. But there are n possibilities for w and so there are n possibilities for wz_0 (one of which is z_0 because one of the values of w is 1). So knowing the solutions of $w^n = 1$ (which is easy) and just one solution of $z^n = a$ (which is not too bad) we can generate n solutions of $z^n = a$. The question is: "Have we now got them all?". It would be most surprising if we had not, but we still have to check—or rather *you* have to check, because that is Problem 8 in Section 1.4.

In solving the equation $z^n = a$ in the set of complex numbers we are extending the notion of roots of a real number and so we make the following definition.

Definition

The solutions of $z^n = a$ are called nth **roots** of a.

We have proved the following theorem (taking Problem 8 of Section 1.4 for granted).

Theorem 1

Every non-zero complex number has exactly n distinct nth roots.

We find the n nth roots of $a \neq 0$ by using the following technique.

Technique

If $a = |a|(\cos\alpha + i\sin\alpha)$ then one solution of $z^n = a$ is

$$z_0 = |a|^{1/n}\left(\cos\frac{\alpha}{n} + i\sin\frac{\alpha}{n}\right)$$

and all the others are obtained by multiplying z_0 in turn by each of the nth roots of $w^n = 1$.

Now the effect of multiplying each of the n nth roots of 1 by z_0 will be to multiply each of their moduli (all equal to 1) by $|a|^{1/n}$ and add α/n to each of their arguments. In other words, the nth roots of the complex number a are obtained from the nth roots of 1 by expanding (or contracting) the unit circle by $|a|^{1/n}$ and rotating by $(\text{Arg } a)/n$. The nth roots of a will be equally spaced around a circle of radius $|a|^{1/n}$. The roots are

$$z_k = |a|^{1/n}\left(\cos\frac{\alpha + 2k\pi}{n} + i\sin\frac{\alpha + 2k\pi}{n}\right), \quad k = 1, 2, \ldots, n.$$

Fig. 9 illustrates this for $n = 8$ and $|a| > 1$.

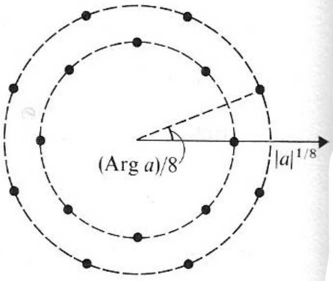


Fig. 9

Example 2

To find the solutions of $z^4 = 1 + i$, the fourth roots of $1 + i$, we proceed as follows.

(i) $1 + i = \sqrt{2}\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)$.

(ii) $(\sqrt{2})^{\frac{1}{4}} = 2^{\frac{1}{8}}$; $\frac{\pi/4}{4} = \frac{\pi}{16}$.

(iii) Thus one root is $2^{\frac{1}{8}}\left(\cos\frac{\pi}{16} + i\sin\frac{\pi}{16}\right)$ and the others can be found because the four of them are equally spaced around the circle of radius $2^{\frac{1}{8}}$ (Fig. 10).

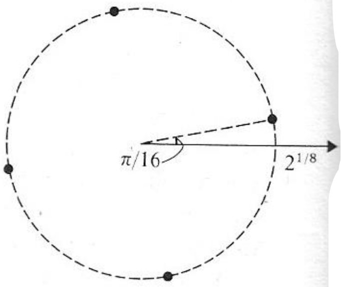


Fig. 10

Summary

In this section we have seen a correspondence between complex numbers and points in the plane using Cartesian coordinates; the real part of the complex number is represented on the real axis ("horizontal") and the imaginary part on the imaginary axis ("vertical"). We can also use polar coordinates (r, θ) in which case r, the modulus of a non-zero complex number z is positive and θ is an argument of z, defined only to within an additive integer multiple of 2π. The principal value of the argument is that value of θ in $(-\pi, \pi]$, and is denoted by Arg z.

Multiplication and division are particularly easy operations to carry out when the numbers are in polar form. We discussed the conjugate of a complex number and how to find the nth roots of a complex number.

Examples

3. (i) If $z = i$, then $|z| = 1$ and $\operatorname{Arg} z = \frac{\pi}{2}$.

   (ii) If $z = -1$, then $|z| = 1$ and $\operatorname{Arg} z = \pi$.

   (iii) If $z = \sqrt{3} - i$ then $|z| = (3 + 1)^{\frac{1}{2}} = 2$ and $\operatorname{Arg} z = -\frac{\pi}{6}$.

   It is usually a good idea to scribble a little diagram like Fig. 11 to make sure that you get the correct angle for Arg z.

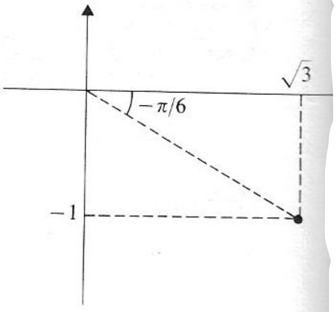


Fig. 11

   (iv) If $z = \sin\alpha - i\cos\alpha$, $|z| = (\cos^2\alpha + \sin^2\alpha)^{\frac{1}{2}} = 1$ and $\operatorname{Arg} z = \alpha - \frac{\pi}{2}$, provided $-\pi/2 < \alpha \leqslant 3\pi/2$.

   In Fig. 12 we have taken $\alpha \in (0, \pi/2)$.

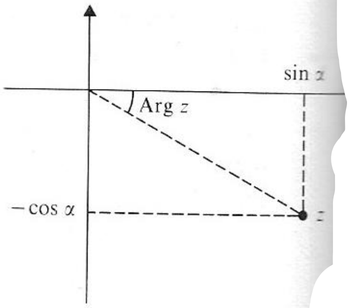


Fig. 12

4. (i) If $|z| = 3$ and $\operatorname{Arg} z = \frac{2\pi}{3}$, then

$$\begin{aligned} z &= 3\cos\frac{2\pi}{3} + i\,3\sin\frac{2\pi}{3} \\ &= -\frac{3}{2} + i\frac{3\sqrt{3}}{2} \quad \text{(Fig. 13).} \end{aligned}$$

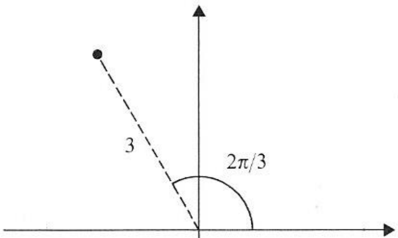


Fig. 13

(ii) If $|z| = 5$ and $\operatorname{Arg} z = -\frac{5\pi}{6}$, then

$$\begin{aligned} z &= 5\cos\left(\frac{-5\pi}{6}\right) + 5\sin\left(\frac{-5\pi}{6}\right) \\ &= \frac{-5\sqrt{3}}{2} - i\frac{5}{2} \quad \text{(Fig. 14).} \end{aligned}$$

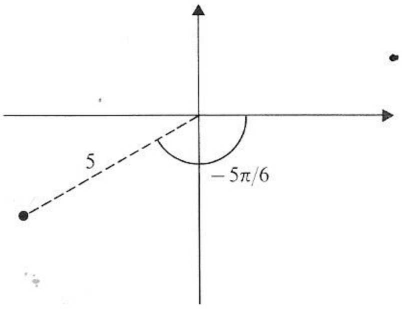


Fig. 14

5. (i) $$\frac{1}{2-3i} = \frac{1}{(2-3i)}\frac{(2+3i)}{(2+3i)} = \frac{2+3i}{4+9} = \frac{2}{13} + \frac{3i}{13}.$$

(ii) $$\frac{3+i}{5-i} = \frac{(3+i)(5+i)}{(5-i)(5+i)} = \frac{14+8i}{26} = \frac{7}{13} + \frac{4i}{13}.$$

6. $$\begin{aligned} (\cos\alpha - i\sin\alpha)^7 &= [\cos\alpha + i\sin(-\alpha)]^7 \\ &= [\cos(-\alpha) + i\sin(-\alpha)]^7 \\ &= \cos(-7\alpha) + i\sin(-7\alpha), \text{ by adding arguments,} \\ &= \cos 7\alpha - i\sin 7\alpha. \end{aligned}$$

Self-Assessment Questions

1. Find $|z|$ and $\operatorname{Arg} z$ in each of the following cases.
   - (i) $z = -i$.
   - (ii) $z = -1 + i$.
   - (iii) $z = -\sqrt{3} - i$.
   - (iv) $z = \cos\alpha + i\sin\alpha, \quad \alpha \in (-\pi, \pi]$.
   - (v) $z = \cos\alpha - i\sin\alpha, \quad \alpha \in (-\pi, \pi]$.
2. Express each of the following in the form $a + ib$ where a and b are real.
   - (i) $\frac{1}{3+2i}$.
   - (ii) $\frac{1}{6+4i}$.
   - (iii) $\frac{1+2i}{1-2i}$.
   - (iv) $\frac{1}{\cos\theta + i\sin\theta}$.
   - (v) $(\cos\theta + i\sin\theta)^{10}$.
   - (vi) $\left(\frac{1}{\cos\theta + i\sin\theta}\right)^{10}$.

Solutions

1. (i) $|z| = 1$, $\operatorname{Arg} z = -\dfrac{\pi}{2}$.

   (ii) $|z| = \sqrt{2}$, $\operatorname{Arg} z = \dfrac{3\pi}{4}$.

   (iii) $|z| = 2$, $\operatorname{Arg} z = -\dfrac{5\pi}{6}$.

   (iv) $|z| = 1$, $\operatorname{Arg} z = \alpha$.

   (v) $|z| = 1$, $\operatorname{Arg} z = -\alpha$, unless $\alpha = \pi$ in which case $\operatorname{Arg} z = \pi$.

2. (i) $\dfrac{3}{13} - \dfrac{2}{13}i$.

   (ii) $\dfrac{1}{2}\left(\dfrac{3}{13} - \dfrac{2}{13}i\right) = \dfrac{3}{26} - \dfrac{1}{13}i$.

   (iii) $\dfrac{(1 + 2i)^2}{(1 - 2i)(1 + 2i)} = -\dfrac{3}{5} + \dfrac{4i}{5}$.

   (iv) $\cos\theta - i\sin\theta$.

   (v) $\cos 10\theta + i\sin 10\theta$.

   (vi) $\cos 10\theta - i\sin 10\theta$.

1.4 PROBLEMS

The solutions immediately follow the last problem.

1. Find the modulus (r) and *an* argument (θ) for each of the following numbers.
   (i) $1 + i$.
   (ii) $(1 + i)^5$.
   (iii) $1 - i$.
   (iv) $1 + i\sqrt{3}$.
   (v) $(1 - i)(1 + i\sqrt{3})$.
   (vi) $-2 - 2i$.
   (vii) $\dfrac{1 - i}{1 + i\sqrt{3}}$.

2. Express each of the following numbers in the form $a + ib$, where a and b are real numbers.

   (i) $\dfrac{6 + i}{7 + 3i}$. (ii) $\dfrac{1 - i}{1 + i}$. (iii) $\dfrac{i + 4}{-4 + i}$.

3. Show that for any complex numbers z and w

   (i) $\operatorname{Re} z = \dfrac{z + \bar{z}}{2}, \quad \operatorname{Im} z = \dfrac{z - \bar{z}}{2i};$

   (ii) $\overline{z + w} = \bar{z} + \bar{w}, \quad \overline{zw} = \bar{z}\,\bar{w}, \quad |\bar{z}| = |z|, \quad \overline{(\bar{z})} = z.$

   These properties are important and will be needed frequently throughout the course.

4. Illustrate the following sets of points of the complex plane where a, b, c, d, α and β are fixed real numbers with $b > a$, $d > c$, $-\pi < \alpha < \beta < \pi$.
   (i) $\{z : \operatorname{Im} z > c\}$.
   (ii) $\{z : \operatorname{Re} z < b\}$.
   (iii) $\{z : c < \operatorname{Im} z < d\}$.
   (iv) $\{z : a < \operatorname{Re} z < b\}$.
   (v) $\{z : a < \operatorname{Re} z < b\} \cap \{z : c < \operatorname{Im} z < d\}$.
   (vi) $\left\{z : \operatorname{Re}\left(\dfrac{z + 3}{z - 2}\right) = 0\right\}$.
   (vii) $\{z : \alpha < \operatorname{Arg} z < \beta\}$.
   (viii) $\{z : z = |z|(\cos\theta + i\sin\theta),\ \theta \in [0, \alpha]\}$.
   (ix) $\{z : 3\operatorname{Re} z + 2\operatorname{Im} z < 6\}$.
   (x) $\{z : 3\operatorname{Re} z + 2\operatorname{Im} z < 6\} \cap \{z : 3\operatorname{Re} z + \operatorname{Im} z > 3\}$
   $\cap \{z : \operatorname{Re} z + \operatorname{Im} z > 2\}$.

5. If n is a multiple of 4, show that the coefficients, c_i in Problem 4 of Section 1.2 satisfy

   $$c_1 - c_3 + c_5 - c_7 + \cdots = 0,$$

   and

   $$1 - c_2 + c_4 - c_6 + \cdots = \pm 2^{n/2}.$$

   Identify the cases when the positive sign should be taken on the right-hand side in the second case. (There is a hint after the solutions.)

6. Let $p(z)$ be a polynomial with real coefficients,
   (i) Prove that $p(\bar{z}) = \overline{p(z)}$.
   (ii) If α is a root of $p(z)$, show that $\bar{\alpha}$ is also a root.

7. (i) Prove by mathematical induction that if n is any positive integer, then

$$|z^n| = |z|^n.$$

   (ii) Show that if n is any positive integer, then

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta.$$

   This result is known as *de Moivre's Theorem**.

   (iii) Show that if n is any positive integer, then

$$(\cos\theta + i\sin\theta)^{-n} = \cos(-n\theta) + i\sin(-n\theta).$$

   This extends de Moivre's Theorem to any integer n.

   (iv) We have seen in the text that de Moivre's Theorem extends to the form: if m is a positive integer, *one of the values* of $(\cos\theta + i\sin\theta)^{1/m}$ is $\cos(\theta/m) + i\sin(\theta/m)$. Extend de Moivre's Theorem to any *rational* index.

8. Show that if z_0 is a solution of $z^n = a$, where $a \neq 0$, then any other solution must be of the form $z_0 w$ where w is a solution of $w^n = 1$.

9. (i) Find the fifth roots of -32.
   (ii) Find the fourth roots of -16.

10. (i) Solve the equation $z^4 + 4z^2 + 8 = 0$.
   (ii) Solve the equation $z^4 + 4iz^2 + 8 = 0$.

Solutions

1. (i) $r = \sqrt{2}, \quad \theta = \frac{\pi}{4}$.

   (We asked you only for *an* argument; you could have given any one of $\pi/4 + 2k\pi$, where k is an integer. An alternative way of expressing this is to write $\theta = \pi/4 \pmod{2\pi}$. Similar remarks apply to the other parts.)

   (ii) $r = (\sqrt{2})^5, \quad \theta = -\frac{3\pi}{4}\left(= \frac{5\pi}{4} \pmod{2\pi}\right)$.

   (iii) $r = \sqrt{2}, \quad \theta = -\frac{\pi}{4}$.

   (iv) $r = 2, \quad \theta = \frac{\pi}{3}$.

   (v) $r = \sqrt{2} \times 2, \quad \theta = -\frac{\pi}{4} + \frac{\pi}{3} = \frac{\pi}{12}$, using the answers to (iii) and (iv).

   (vi) $r = \sqrt{8}, \quad \theta = -\frac{3\pi}{4}$.

   (vii) $r = \frac{\sqrt{2}}{2}, \quad \theta = -\frac{\pi}{4} - \frac{\pi}{3} = -\frac{7\pi}{12}$, using the answers to (iii) and (iv).

---

* *Abraham de Moivre* (1667–1754) was born in France, but came to England when still a small boy. His interest in Mathematics started when he came across a copy of Newton's *Principia* by chance. He is best known for the theorem stated above, and for calculating the various quadratic factors of $x^{2n} - 2px^n + 1$, and is also remembered as one of the founders of probability theory.

2. (i) $\dfrac{(6+i)(7-3i)}{7^2+3^2} = \dfrac{45}{58} - \dfrac{11i}{58}$.

   (ii) $\dfrac{(1-i)(1-i)}{1^2+1^2} = -i$.

   (iii) $\dfrac{(i+4)(-4-i)}{4^2+1^2} = \dfrac{-15}{17} - \dfrac{8i}{17}$.

3. (i) $\dfrac{z+\bar{z}}{2} = \dfrac{x+iy+x-iy}{2} = x = \operatorname{Re} z;$

   $\dfrac{z-\bar{z}}{2i} = \dfrac{x+iy-x+iy}{2i} = y = \operatorname{Im} z.$

   (Note that we have taken $z = x + iy$; since this is the standard representation, we do not normally state it explicitly.)

   (ii) If $w = u + iv$, then

$$\begin{aligned}
\overline{z+w} &= \overline{x+iy+u+iv}\\
&= x+u-i(y+v)\\
&= x-iy+u-iv\\
&= \bar{z}+\bar{w};\\
\overline{zw} &= \overline{(x+iy)(u+iv)}\\
&= \overline{xu-yv+i(yu+xv)}\\
&= xu-yv-i(yu+xv)\\
&= (x-iy)(u-iv)\\
&= \bar{z}\bar{w};\\
|\bar{z}| &= \sqrt{x^2+(-y)^2} = \sqrt{x^2+y^2} = |z|;\\
\bar{z} &= x-iy, \text{ so } \overline{(\bar{z})} = \overline{x-iy} = x+iy = z.
\end{aligned}$$

4. (i)

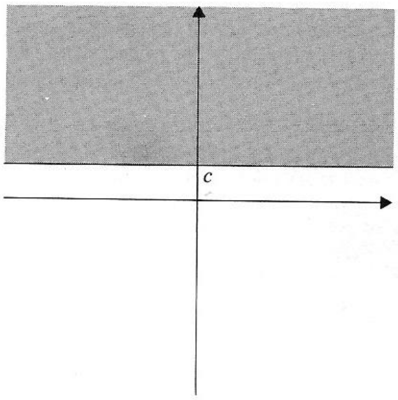


Fig. 15

(ii)

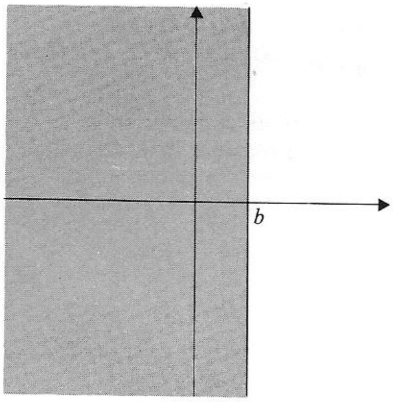


Fig. 16

(iii)

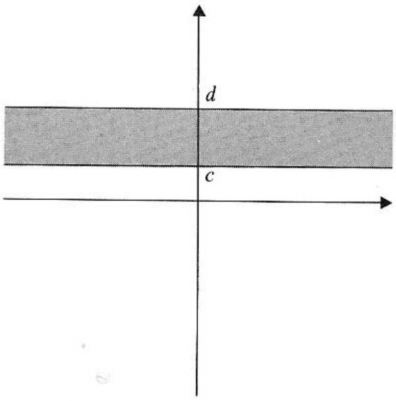


Fig. 17

(iv)

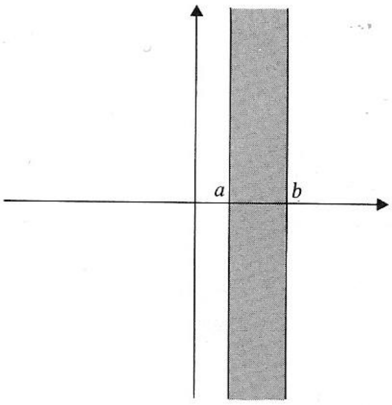


Fig. 18

(v)

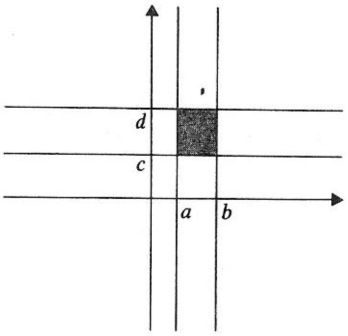


Fig. 19

(vi) $\operatorname{Re}\left(\dfrac{z+3}{z-2}\right) = \operatorname{Re}\left(\dfrac{x+3+iy}{x-2+iy}\right)$

$$= \operatorname{Re}\left(\frac{(x+3+iy)(x-2-iy)}{(x-2)^2+y^2}\right)$$

and this is zero if and only if

$$(x+3)(x-2)+y^2=0.$$

Thus x and y satisfy

$$x^2+x-6+y^2=0,$$

which is equivalent to

$$(x+\tfrac{1}{2})^2+y^2=6\tfrac{1}{4}.$$

The locus is a circle with centre $(-\frac{1}{2}+0i)$ and radius $\frac{5}{2}$ (Fig. 20). The point $z = 2$ on this circle must be excluded from the locus, because $(z+3)/(z-2)$ is not defined for $z = 2$.

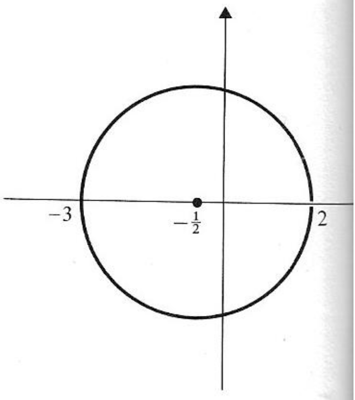


Fig. 20

When you are used to manipulating complex numbers you will be able to work more elegantly as follows. $\operatorname{Re}\left(\dfrac{z+3}{z-2}\right)=0$ gives $\operatorname{Re}\big((z+3)(\bar{z}-2)\big)=0$, since $(z-2)(\bar{z}-2)$ is real. Thus

$$\operatorname{Re}(z\bar{z}-2z+3\bar{z}-6)=0,$$

and so

$$x^2+y^2-2x+3x-6=0.$$

(vii)

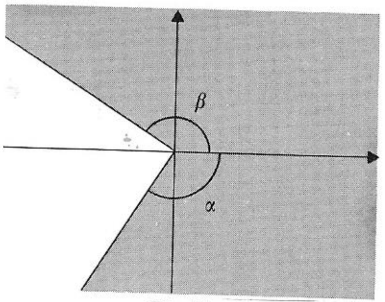


Fig. 21

(viii)

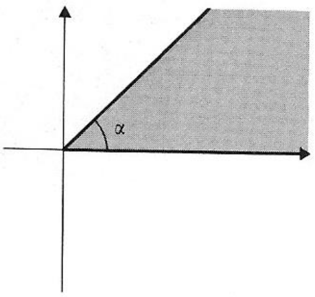


Fig. 22

(ix) $3\operatorname{Re} z + 2\operatorname{Im} z < 6$ corresponds to $3x+2y<6$ and so the set is as indicated in Fig. 23.

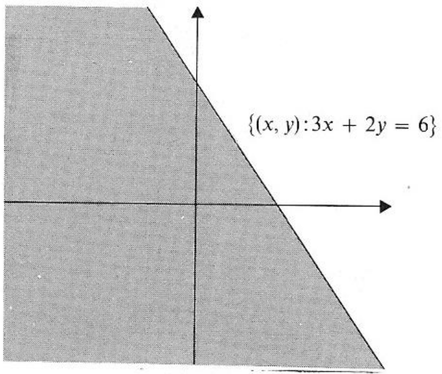


Fig. 23

(x)

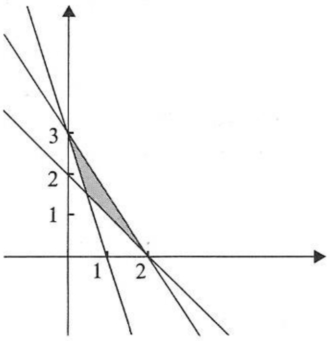


Fig. 24

(The regions in Figs. 15, 16 and 23 are referred to as "half-planes". Figs. 17 and 18 illustrate "strips"; Fig. 19 shows a "rectangle", and Fig. 24 a "triangle". Note also that in Figs. 15 to 19 we have illustrated only the case a, b, c and d positive. In Fig. 21 we have chosen $\alpha < -\pi/2$ and $\beta > \pi/2$. In Fig. 22 we have chosen $0 < \alpha < \pi/2$.)

Style for figures: It is difficult to develop a completely consistent style for figures. However, you should find the following guidelines useful. (i) Infinite (unbounded) sets are shown by squared-off shading as in Figs. 15, 16, 21, 22 and 23. (ii) Sets which do not include their "boundary" lines have thin boundary lines (for example, Figs. 15, 16, 17, 18, 21, 23 and 24). (iii) Thick boundary lines indicate that the set includes its boundary lines. (iv) Most other conventions are self-evident or obvious from the context.

5. Substituting $z = i$ in $(1 + z)^n = 1 + c_1 z + c_2 z^2 + \cdots + c_n z^n$ gives

$$(1 + i)^n = (1 - c_2 + c_4 - c_6 + \cdots) + i(c_1 - c_3 + c_5 - \cdots).$$

Since $\operatorname{Arg}(1 + i) = \pi/4$, $(1 + i)^n$ is real if n is a multiple of 4, and so equating real and imaginary parts gives

$$\begin{aligned} c_1 - c_3 + c_5 - c_7 + \cdots &= 0 \\ 1 - c_2 + c_4 - c_6 + \cdots &= (1 + i)^n. \end{aligned}$$

Since $|1 + i| = \sqrt{2}$, we have $|(1 + i)^n| = 2^{n/2}$.

We have seen that if n is a multiple of 4 then $(1 + i)^n$ is real, that is $\operatorname{Arg}(1 + i)^n$ is either 0 (in which case $(1 + i)^n$ is positive) or π (in which case $(1 + i)^n$ is negative). But, since an argument of $1 + i$ is $\pi/4$, an argument of $(1 + i)^n$ is $n\pi/4$.

If n is an even multiple of 4, that is $n = 2k \times 4$ for some integer k, an argument of $(1 + i)^n$ is $2k\pi$. In other words, $\operatorname{Arg}(1 + i)^n = 0$ and so $(1 + i)^n$ is positive.
If n is an odd multiple of 4, that is $n = (2k + 1) \times 4$ for some integer k, an argument of $(1 + i)^n$ is $(2k + 1)\pi$. In other words, $\operatorname{Arg}(1 + i)^n = \pi$ and so $(1 + i)^n$ is negative.

6. (i) $p(\bar{z}) = c_0 + c_1\bar{z} + c_2(\bar{z})^2 + c_3(\bar{z})^3 + \cdots + c_n(\bar{z})^n$. By an extension of a result in Problem 3(ii), $\bar{z}^k = \overline{z^k}$, $k = 1, \ldots, n$. Also, if c_k is real, $c_k\bar{w} = \overline{c_k w}$, and again from Problem 3(ii), the conjugate of a sum is the sum of the conjugates and so $p(\bar{z}) = \overline{p(z)}$.

   (ii) If α is a root of the polynomial then $p(\alpha) = 0$ and so $\overline{p(\alpha)} = 0$ and thus from (i), $p(\bar{\alpha}) = 0$; in other words, $\bar{\alpha}$ is also a root.

7. (i) The result is trivially true for $n = 1$.
   If the result is true for $n = k$, that is $|z^k| = |z|^k$, then

$$\begin{aligned} |z^{k+1}| &= |z^k z| \\ &= |z^k| \cdot |z| \quad \text{(because } |zw| = |z| \cdot |w|) \\ &= |z|^k \cdot |z| \quad \text{(by hypothesis)} \\ &= |z|^{k+1}. \end{aligned}$$

   The result follows by mathematical induction.

   (ii) The result is trivially true for $n = 1$. If the result is true for $n = k$, then

$$\begin{aligned} (\cos\theta + i\sin\theta)^{k+1} &= (\cos\theta + i\sin\theta)^k(\cos\theta + i\sin\theta) \\ &= [\cos(k\theta) + i\sin(k\theta)](\cos\theta + i\sin\theta) \\ &= \cos(k + 1)\theta + i\sin(k + 1)\theta, \end{aligned}$$

   from the ordinary properties of multiplication of two complex numbers.
   The result follows by mathematical induction.

(iii) This follows from (ii) by noting that

$$\begin{aligned}(\cos\theta + i\sin\theta)^{-1} &= \frac{1}{\cos\theta + i\sin\theta}\\ &= \frac{\cos\theta - i\sin\theta}{\cos^2\theta + \sin^2\theta}\\ &= \cos(-\theta) + i\sin(-\theta).\end{aligned}$$

Hence

$$\begin{aligned}(\cos\theta + i\sin\theta)^{-n} &= [(\cos\theta + i\sin\theta)^{-1}]^n\\ &= [\cos(-\theta) + i\sin(-\theta)]^n\\ &= \cos(-n\theta) + i\sin(-n\theta).\end{aligned}$$

(iv) Let r be any rational number. Then $r = n/m$ where n is an integer and m is a positive integer. Since m is a positive integer, one of the values of $(\cos\theta + i\sin\theta)^{1/m}$ is $\cos(\theta/m) + i\sin(\theta/m)$, and so, from (ii) and (iii), one of the values of $(\cos\theta + i\sin\theta)^{n/m}$ is $\cos(n\theta/m) + i\sin(n\theta/m)$.

8. Since $a \neq 0$, we have $z_0 \neq 0$ and so any complex number z can be written in the form z_0w, just by choosing $w = \dfrac{z}{z_0}$. If z is a solution of $z^n = a$, then

$$(z_0w)^n = a$$

and so

$$z_0^n w^n = a.$$

But $z_0^n = a$ and so we must have $w^n = 1$.

9. (i) Since $-32 = 32[\cos(-\pi) + i\sin(-\pi)]$, one fifth root of -32 is

$$2\left[\cos\left(\frac{-\pi}{5}\right) + i\sin\left(\frac{-\pi}{5}\right)\right].$$

Since the five fifth roots of -32 are equally spaced around the circle with centre at the origin and radius 2, the other roots are

$$2\cos\left(\frac{-\pi + 2k\pi}{5}\right) + i\sin\left(\frac{-\pi + 2k\pi}{5}\right), \quad k = 1, 2, 3, 4.$$

(Note that $k = 3$ gives the real fifth root -2. The above method is the systematic one; we could, however, have made use of -2 as the one obvious fifth root of -32 from the start.)

(ii) Since $-16 = 16[\cos(-\pi) + i\sin(-\pi)]$, one fourth root of -16 is

$$2\left[\cos\left(\frac{-\pi}{4}\right) + i\sin\left(\frac{-\pi}{4}\right)\right].$$

Since the four fourth of -16 are equally spaced around the circle with centre at the origin and radius 2, the other roots are

$$2\left[\cos\left(\frac{-\pi + 2k\pi}{4}\right) + i\sin\left(\frac{-\pi + 2k\pi}{4}\right)\right], \quad k = 1, 2, 3.$$

10. (i) Putting $w = z^2$ gives

$$w^2 + 4w + 8 = 0,$$

and so

$$w = \frac{-4 \pm \sqrt{16 - 32}}{2} = -2 \pm 2i.$$

So z can take the values of the two square roots of $-2 + 2i$ and the two square roots of $-2 - 2i$. Since

$$-2 + 2i = \sqrt{8}\left(\cos\frac{3\pi}{4} + i\sin\frac{3\pi}{4}\right),$$

the two square roots of $-2 + 2i$ are

$$8^{\frac{1}{4}}\left(\cos\frac{3\pi}{8} + i\sin\frac{3\pi}{8}\right)$$

and

$$8^{\frac{1}{4}}\left[\cos\left(\frac{3\pi}{8} + \pi\right) + i\sin\left(\frac{3\pi}{8} + \pi\right)\right] = -8^{\frac{1}{4}}\left(\cos\frac{3\pi}{8} + i\sin\frac{3\pi}{8}\right).$$

Similarly, we can show that the two square roots of $-2-2i$ are

$$\pm 8^{\frac{1}{4}}\left(\cos\frac{3\pi}{8} - i\sin\frac{3\pi}{8}\right).$$

(Alternatively, we could use the result of Problem 6(ii) to find the last two roots, since $z^4 + 4z^2 + 8$ is a polynomial with real coefficients.)

(ii) Putting $w = z^2$ gives

$$w^2 + 4iw + 8 = 0,$$

and so

$$w = \frac{-4i \pm \sqrt{-16-32}}{2},$$

that is

$$w = (\sqrt{12}-2)i \quad \text{or} \quad w = -(\sqrt{12}+2)i.$$

Since $(\sqrt{12}-2)i = (\sqrt{12}-2)\left(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}\right)$, the two square roots of $(\sqrt{12}-2)i$ are

$$z = (\sqrt{12}-2)^{\frac{1}{2}}\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)$$

and

$$\begin{aligned} z &= (\sqrt{12}-2)^{\frac{1}{2}}\left[\cos\left(\frac{\pi}{4}+\pi\right) + i\sin\left(\frac{\pi}{4}+\pi\right)\right] \\ &= (\sqrt{12}-2)^{\frac{1}{2}}\left[\cos\left(\frac{-3\pi}{4}\right) + i\sin\left(\frac{-3\pi}{4}\right)\right]. \end{aligned}$$

Since

$$-(\sqrt{12}+2)i = (\sqrt{12}+2)\left[\cos\left(-\frac{\pi}{2}\right) + i\sin\left(-\frac{\pi}{2}\right)\right],$$

the two square roots of $-(\sqrt{12}+2)i$ are

$$z = (\sqrt{12}+2)^{\frac{1}{2}}\left[\cos\left(\frac{-\pi}{4}\right) + i\sin\left(\frac{-\pi}{4}\right)\right]$$

and

$$\begin{aligned} z &= (\sqrt{12}+2)^{\frac{1}{2}}\left[\cos\left(\frac{-\pi}{4}+\pi\right) + i\sin\left(\frac{-\pi}{4}+\pi\right)\right] \\ &= (\sqrt{12}+2)^{\frac{1}{2}}\left[\cos\left(\frac{3\pi}{4}\right) + i\sin\left(\frac{3\pi}{4}\right)\right]. \end{aligned}$$

Note that the four roots of $z^4 + 4iz^2 + 8 = 0$ do not occur in conjugate pairs: the coefficient of z^2 is not real.

Hint

5. Try substituting a suitable value for z in

$$(1+z)^n = 1 + c_1 z + c_2 z^2 + \cdots + c_n z^n.$$

1.5 SETS OF POINTS IN THE COMPLEX PLANE

We have seen in Problem 4 of Section 1.4 how some sets of points of the complex plane can be described algebraically in terms of operations on **C**. We now use the modulus function to take this a step further by defining *discs* in the complex plane. As we shall see, discs are extensively used in arguments involving limits. Another type of region that is useful in analytic arguments is a rectangular region, and so we conclude this section with a discussion of the Nested Rectangles Theorem.

* * * * * * * *

You will no doubt recall that in real analysis extensive use is made of the modulus function $x \longrightarrow |x|$ $(x \in \mathbf{R})$. It gives us a way of measuring the "closeness" of two numbers, which we exploit in writing expressions such as

$$\text{"}\ldots |f(x) - l| < \varepsilon \text{ for all } x \text{ such that } 0 < |x - a| < \delta \ldots\text{"}.$$

We are going to need a similar device in complex analysis and, of course, it is again provided by the modulus function, this time with domain **C**. In Section 1.3 we defined $|z|$, the modulus of the complex number z. The **modulus function** is the function which maps a complex number to its modulus:

$$z \longrightarrow |z|, \qquad z \in \mathbf{C}.$$

Note that the modulus function maps complex numbers to real numbers.

The modulus function with domain **C** is just an extension of the modulus function that we use in real analysis in the sense that if we restrict its domain to **R** then the two functions are equal. The value at z of the modulus function, $|z|$, tells us the distance of the point z from the origin. Similarly, just as $|a - b|$ is the distance between the two real numbers a and b, so $|z - w|$ is the **distance between the two complex numbers** z and w.

More explicitly, if $z = x + iy$ and $w = u + iv$, then

$$\begin{aligned} |z - w| &= |(x - u) + i(y - v)| \\ &= \sqrt{(x - u)^2 + (y - v)^2}, \end{aligned}$$

which is the usual (Euclidean) distance between the points (x, y) and (u, v).

The function $d:(z, w) \longrightarrow |z - w|$, $(z, w) \in \mathbf{C} \times \mathbf{C}$, is a *distance function* in the usual sense of the word in that it turns the set of complex numbers into a metric space (see Problem 7 in Section 1.6 and *Unit M231 2, Functions and Graphs,* Section 2.7). Among other things this means that we have the **triangle inequality**:

$$|z - w| \leqslant |z - v| + |v - w|, \quad z, w, v \in \mathbf{C}.$$

(Note that this is an inequality between real numbers (moduli): the modulus function enables us to write inequalities which *involve* complex numbers.)

This inequality reduces to the usual triangle inequality of real analysis when z, w and v are real, but the two-dimensional nature of complex numbers explains why the triangle inequality is so called. You will find as this course proceeds that the triangle inequality is as vital to complex analysis as it is to real analysis: it also yields many more inequalities, some of which are in the next problems section.

In real analysis, the modulus function enables us to describe an interval around a point. For example, we talk of an ε-neighbourhood of a point x_0 as the interval $\{x : |x - x_0| < \varepsilon\}$. The end points of this interval are simply the two elements of the set $\{x : |x - x_0| = \varepsilon\}$. In the complex plane the analogous bounding set is $\{z : |z - z_0| = \varepsilon\}$, which is the set of all points at a distance ε from z_0, that is to say the circle centre z_0, radius ε (Fig. 25).

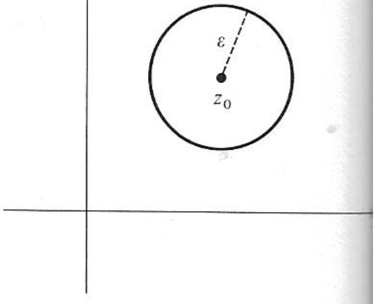


Fig. 25

Thus in complex analysis, the analogue of an interval is the set of points whose boundary is a circle. Such a set is called a *disc*. In particular, we make the following definitions.

Definition

The set $\{z : |z - z_0| < k, k > 0\}$ is an **open disc**.
The set $\{z : |z - z_0| \leqslant k, k > 0\}$ is a **closed disc**.

If we are not sure whether a disc is open or closed, or if we do not want to be specific, or if some but not all points of the bounding circle are included we shall refer simply to the "disc". We shall also sometimes drop the set notation and write, for example, "the closed disc $|z - z_0| \leqslant k$" when there is no likelihood of misunderstanding. One other point of convention: we shall refer to the disc centre the origin and radius 1 as the (open or closed) *unit disc*.

Two concentric circles specify another type of set of complex numbers—the set of points between the two circles.

Definition

The set $\{z : a < |z - z_0| < b, 0 < a < b\}$ is an **open annulus.**
The set $\{z : a \leqslant |z - z_0| \leqslant b, 0 < a \leqslant b\}$ is a **closed annulus.**

If we do not wish to be specific, or if some but not all of the points of the bounding circles are included, we drop the adjective "open" or "closed". As for discs, we often write, for example, "the open annulus $a < |z - z_0| < b$".

Another term which we shall find useful is "bounded set of points".

Definition

A set of points is said to be **bounded** if all points of the set lie in a closed disc with centre the origin.

As an example of the way we can use the modulus function in complex analysis, consider how we can define the limit of a sequence of complex numbers. In real analysis, an infinite sequence is simply a function from the set of natural numbers, **N**, to **R**; in complex analysis an infinite sequence is a function from **N** to **C**. Just as the terms of a real sequence can be plotted on the number line, so the terms of a complex sequence can be plotted on the complex plane. Convergence is defined precisely as for real sequences.

A sequence $\{z_n\}$ **converges to** l if for every $\varepsilon > 0$ there is a natural number N such that, for all natural numbers n,

if $n > N$, then $|z_n - l| < \varepsilon$.

The number l is called the **limit** of the sequence.

In other words, for all $n > N$ every term in the sequence is in the (open) disc centre l radius ε. (Later in the course we shall study such limits and the attendant theorems.)

Notice that the analytic form of the definition looks just like that of real analysis —$\{z_n\}$ converges to l if and only if $\lim_{n \to \infty} |z_n - l| = 0$ —but the geometric interpretation is much more interesting. If the values of ε in this definition are made successively smaller, taking values $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots$ then we get a succession of discs $|z - l| < \varepsilon_1$, $|z - l| < \varepsilon_2$, and so on, each disc being a subset of its predecessor and containing its successor. Sequences of sets with this property are often used in complex analysis; they are called **nested sequences**. They are used to give a method of proof similar to the repeated bisection arguments of real analysis.

Suppose S_1 and S_2 are two nonempty sets, with $S_1 \supseteq S_2$. Then there is at least one point common to both sets. Similarly, if $S_1, S_2, \ldots, S_n$ are n nonempty sets with

$$S_1 \supseteq S_2 \supseteq S_3 \supseteq \cdots \supseteq S_n$$

—that is to say the sets form a nested sequence—then there is at least one point in common to all n sets. If we try to extend this idea to an infinite sequence of sets then a number of snags can arise which *may* result in the intersection of all the sets being empty.

(i) For example, taking sets in **R**, suppose S_n is the open interval $(0, 1/n)$. In this case the sets are nested, for $S_n \supset S_{n+1}$ for all n, but there is no point in common to every set—the obvious candidate, 0, has been excluded because the intervals are open.

(ii) In other circumstances the "obvious" candidate might be excluded not because the intervals are open but because the sets are subsets of a set which is not complete—**Q**, the set of rationals, for example (see *Unit M231 14, Sequences and Series*). The sets given by $S_n = \{x : |x - \sqrt{2}| < 1/n, x \in \mathbf{Q}\}$ are nested but there is no number common to them all. The number $\sqrt{2}$ is excluded because $\sqrt{2} \notin \mathbf{Q}$.

(iii) The third snag that might arise with nested sets of real numbers concerns sets such as intervals $[n, \infty)$. These sets are not open intervals and so are not subject to the first objection. They are also subsets of the complete set, **R**, and so are not subject to the second objection. But there is no real number in every S_n.

In real analysis, we get over these difficulties by using nested sets that are closed intervals (and therefore subsets of the complete set **R**) and that are such that the lengths of the intervals have limit 0. There are several ways of extending this idea to **C** and indeed to metric spaces in general, but a simple and useful way is to extend the idea of an interval to that of a rectangle.

Definition

The set $\{z : z \in \mathbf{C}, a \leqslant \operatorname{Re} z \leqslant b, c \leqslant \operatorname{Im} z \leqslant d, a, b, c, d \in \mathbf{R}\}$ is a **closed rectangle.**
The set $\{z : z \in \mathbf{C}, a < \operatorname{Re} z < b, c < \operatorname{Im} z < d, a, b, c, d \in \mathbf{R}\}$ is an **open rectangle.**

(Now look back at Problem 4 part (v) of Section 1.4.)

If some but not all of the bounding points are included, we shall refer simply to a *rectangle* and we shall also use this term if we do not wish to be specific.

As a concept analogous to the length of an interval we can use the length of the largest side for a rectangle S.* We denote this by $\ell(S)$, so

$$\ell(S) = \max\big((b - a), (d - c)\big).$$

Before we propose the Nested Rectangles Theorem we introduce the idea of the *intersection of a collection of sets*. Suppose that C is a collection of sets and that these sets may be indexed so that we can write C as

$$C = \{S_i : i \in \mathbf{N}\}.$$

We extend the idea of the intersection of two sets to the intersection of the collection of sets in C by regarding the intersection of the S_i as the set $\{x : x \text{ is in } S_i \text{ for every } i \in \mathbf{N}\}$; we denote this set by $\bigcap_{i \in \mathbf{N}} S_i$. In other words, $\bigcap_{i \in \mathbf{N}} S_i$ denotes the set of elements which are common to all the S_i.

---

* In *Unit 2, Continuous Functions*, we shall use the length of the perimeter of a rectangle.

We can now propose the Nested Rectangles Theorem.

Theorem 2 (The Nested Rectangles Theorem)

Let $\{S_n\}$ be a sequence of nested closed rectangles, with $\lim_{n\to\infty} \ell(S_n) = 0$, then

(i) $$S = \bigcap_{i\in\mathbf{N}} S_i$$

is nonempty and contains a unique complex number z_0, and

(ii) given $\varepsilon > 0$, there is an $N \in \mathbf{N}$ such that

$$S_n \subset \{z : |z - z_0| < \varepsilon\} \quad \text{for all } n > N.$$

(This second part of the theorem tells us that we can always ensure that "eventually" all the rectangles (or, if you prefer, "all but a finite number") lie inside an arbitrarily small disc centred at z_0.)

Proof

(i) Probably the simplest proof of this part is to use the Nested Intervals Theorem from real analysis (see **Spivak**). If S_n is bounded by the lines $\operatorname{Re} z = a_n$, $\operatorname{Re} z = b_n$, $\operatorname{Im} z = c_n$, $\operatorname{Im} z = d_n$, then the sets $I_n = [a_n, b_n]$ form a sequence of nested intervals, and so there is a unique real number a belonging to every I_n. Similarly, the nested intervals $J_n = [c_n, d_n]$ have a unique real number b belonging to all of them. Since $S_n = I_n \times J_n$, the unique number $z_0 = a + ib$ belongs to all the S_n and so S contains z_0.

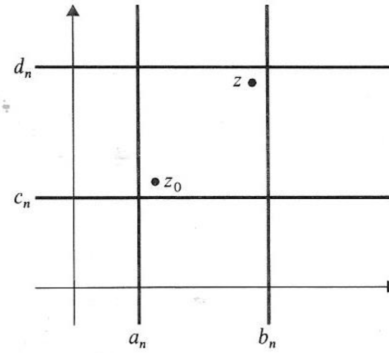


Fig. 26

(ii) Let $z \in S_n$. By part (i) $z_0 \in S_n$.
Now

$$\begin{aligned} |z - z_0|^2 &\leqslant |a_n - b_n|^2 + |c_n - d_n|^2 \\ &\leqslant 2[\ell(S_n)]^2. \end{aligned}$$

But $\lim_{n\to\infty} \ell(S_n) = 0$; therefore, given $\varepsilon > 0$, there is a number N such that

$$\ell(S_n) < \frac{\varepsilon}{\sqrt{2}} \quad \text{for } n > N.$$

Thus, for $z \in S_n$ and $n > N$,

$$\begin{aligned} |z - z_0| &\leqslant \sqrt{2}\,\ell(S_n) \\ &< \varepsilon. \end{aligned}$$

In other words, $S_n \subset \{z : |z - z_0| < \varepsilon\}$. ■

Summary

The modulus function provides us with a measure of distance that turns the set of complex numbers into a metric space in much the same way as does the modulus function defined on **R**. From the point of view of analysis the importance of this is that we can talk of the closeness of two complex numbers. We can then define the limit of a sequence of complex numbers in a way which is almost identical to the definition of the limit of a real sequence. Another analogue of real analysis arises with the Nested Rectangles Theorem. This will play a role in complex analysis similar to the least upper bound property (or, equivalently, the Nested Intervals Theorem) of real analysis.

The modulus function also gives us a way of describing certain sets of points in the complex plane in a convenient algebraic form. These descriptions are the subject of some of the problems in the next section.

Self-Assessment Questions

1. Find the distance between the numbers $2 - i$ and $1 + 3i$.

2. In each of the following cases, classify the statement as True or False.

   (i) The set $\{z : a \leqslant \operatorname{Re} z \leqslant b\} \cap \{z : c < \operatorname{Im} z < d\}$, where $a, b, c, d \in \mathbf{R}$, is a rectangle.

   (ii) The set in (i) is a closed rectangle.

   (iii) The set in (i) is an open rectangle.

   (iv) The set $\{z : |z| \leqslant k, k > 0\}$ is a disc.

   (v) The set in (iv) is a closed disc.

   (vi) The sets S_n, $n = 1, 2, 3, \ldots$, where

   $$S_n = \{z : |z - n| \leqslant 1\}$$

   form a nested sequence.

   (vii) The sets S_n, $n = 1, 2, 3, \ldots$, where

   $$S_n = \{z : |z - 1| \leqslant n\}$$

   form a nested sequence.

   (viii) The sets S_n, $n = 1, 2, 3, \ldots$, where

   $$S_n = \left\{z : |z - 1| < \frac{1}{n}\right\}$$

   form a nested sequence.

   (ix) The sets S_n, $n = 1, 2, 3, \ldots$, where

   $$S_n = \{z : |z| > n\},$$

   form a nested sequence.

Solutions

1. $|(2 - i) - (1 + 3i)| = |1 - 4i| = \sqrt{17}$.

2. (i) True. (ii) False. (iii) False. (iv) True. (v) True.

   (vi) False (Fig. 27). (vii) False (Fig. 28).

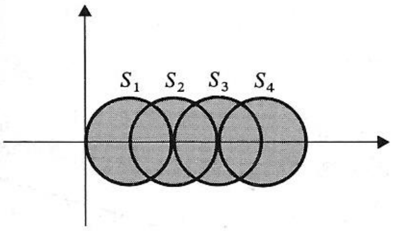
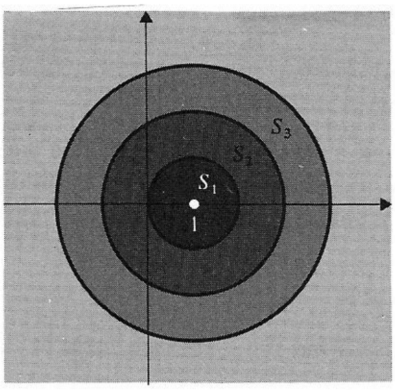


Fig. 27



(viii) True (Fig. 29). (ix) True (Fig. 30).

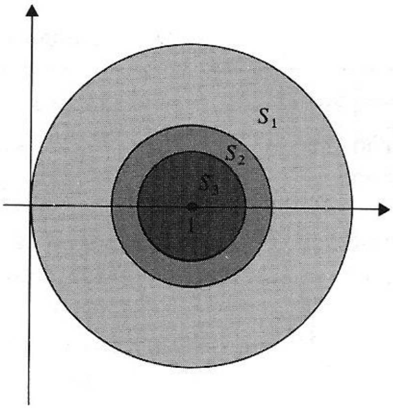


Fig. 29

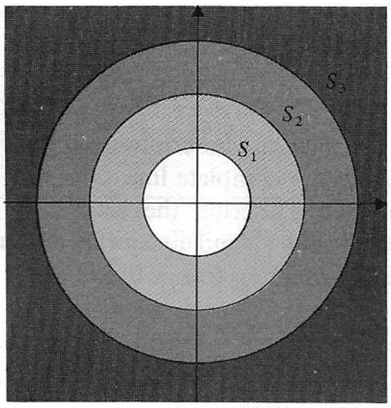


Fig. 30

1.6 PROBLEMS

The solutions immediately follow the last problem.

1. Sketch the curves corresponding to the following equations.

   (i) $|z + 1| = 1$.

   (ii) $|z + 3 - i| = |z - 2|$.

2. Find the Cartesian equation (that is an equation connecting x and y directly) of each of the curves in Problem 1.

3. If you have ever tried to cut a flower bed by tying the ends of a piece of string to two canes so that the string is slack and then fiddling about with another cane keeping the string taut, as in Fig. 31, then you will know that the equation

$$|z - a| + |z - b| = k$$

is the equation of an ellipse. Find the Cartesian form of the equation

$$|z - 1| + |z - 2| = 3.$$

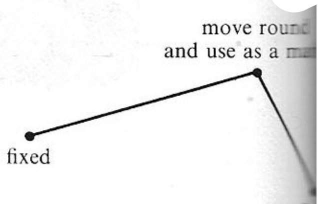


Fig. 31

4. Sketch each of the following sets of points, where $a, b, c, d \in \mathbf{C}$:

   (i) $1 < |z + 1 - i| < 2$;

   (ii) $k_1 \leqslant |z - a| + |z - b| \leqslant k_2$, $|a - b| < k_1 < k_2$;

   (iii) $\left\{z : \left|\dfrac{z - a}{z - b}\right| \leqslant 1\right\} \cap \left\{z : \left|\dfrac{z - c}{z - d}\right| \leqslant 1\right\}$.

5. (i) We sometimes want an algebraic description of a *line segment* in the complex plane, rather than a complete line. One way of doing this is to use a real parameter. Describe the set $\{z : z = tz_1 + (1 - t)z_2, t \in [0, 1]\}$ where t is real and z_1 and z_2 are fixed complex numbers. Identify the points corresponding to $t = 0, \frac{1}{3}, \frac{1}{2}$ and 1 respectively.

   (ii) Describe the set $A = \left\{z : \text{Im}\left(\dfrac{z - z_2}{z_1 - z_2}\right) = 0\right\}$. $\left(\text{Hint: Put } t = \dfrac{z - z_2}{z_1 - z_2}.\right)$

6. (i) Prove the following inequalities.

   (a) $|\text{Re}\, z| \leqslant |z|$.

   (b) $|\text{Im}\, z| \leqslant |z|$.

   (c) $2|\text{Re}\, z| \cdot |\text{Im}\, z| \leqslant |z|^2$.

   (d) $|z| \leqslant |\text{Re}\, z| + |\text{Im}\, z| \leqslant \sqrt{2}|z|$.

   (ii) Show that

$$z\bar{w} + \bar{z}w = 2\,\text{Re}(z\bar{w}).$$

   (iii) Use the formula

$$(z + w)(\bar{z} + \bar{w}) = z\bar{z} + w\bar{w} + z\bar{w} + \bar{z}w$$

   together with part (ii) and a variation of (a) of part (i) to prove the triangle inequality:

$$|z + w| \leqslant |z| + |w|.$$

   (iv) By writing $z = w + (z - w)$, prove that

$$|z| - |w| \leqslant |z - w|.$$

   Prove, similarly, that

$$|w| - |z| \leqslant |w - z|$$

   and hence that

$$\big||z| - |w|\big| \leqslant |z - w|.$$

   Interpret this result geometrically.

7. Show that the function $d: \mathbf{C} \times \mathbf{C} \longrightarrow \mathbf{R}$ where $d(z, w) = |z - w|$ is a distance function; that is to say d satisfies:

   (i) $d(z, w) = d(w, z)$;

   (ii) $d(z, w) \geqslant 0, \quad d(z, w) = 0 \Leftrightarrow z = w$;

   (iii) $d(z, w) \leqslant d(z, a) + d(a, w)$.

   (See *Unit M231 2, Functions and Graphs,* Section 2.7.)

8. Prove that

   (i) $|zw| \geqslant \frac{1}{2}(z\bar{w} + \bar{z}w)$;

   (ii) $|z|^2 + |w|^2 \geqslant z\bar{w} + \bar{z}w$.

9. Find an upper bound for each of the following sets.

   (i) $\left\{\left|\dfrac{1}{z^2 + 1}\right| : |z| = 2\right\}$.

   (ii) $\left\{\left|\dfrac{1}{3 + 4z}\right| : |z| = 1\right\}$.

The next problem deals with the Bolzano–Weierstrass* Theorem. To pose it we require the following definition.

Definition

Let S be any set, and z_0 any point in $\mathbf{C}$. Then z_0 is called a **cluster point** of S if for all $\varepsilon > 0$ there is $z \in S$ such that $0 < |z - z_0| < \varepsilon$. (Note that z_0 need not belong to S.)

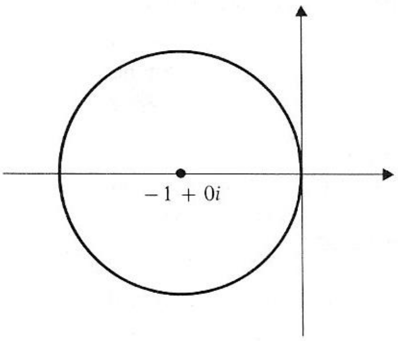
(Note that in *Unit M231 4, Continuous Functions,* we called such a point (in $\mathbf{R}$) an accumulation point.)

---

10. Prove the Bolzano–Weierstrass Theorem: if S is a bounded set and contains infinitely many complex numbers then S has at least one cluster point.

Solutions

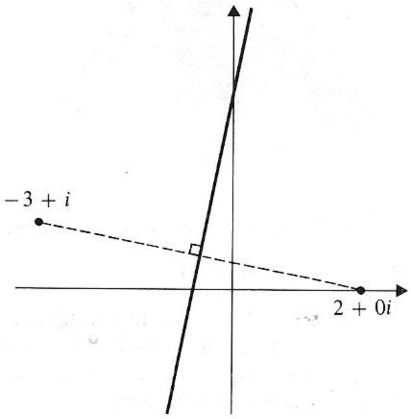
1. (i)



The set of all points at a distance 1 from $-1 + 0i$

Fig. 32

(ii)



The set of all points equidistant from $-3 + i$ and $2 + 0i$

Fig. 33

---

* *Bernard Bolzano* (1781–1848) was professor of the Philosophy of Religion, in Prague. He wrote a treatise entitled *Paradoxes of the Infinite*, but this did not appear until after his death.

Karl Weierstrass (1815–1897) was, together with Cauchy and Riemann, one of the founders of complex analysis. He led an uneventful life, much of it spent in Berlin where he was the professor of Mathematics. He is remembered both for his work on elliptic functions, and also for his attempts to put analysis on a firm logical foundation. You will come across his name in connection with the Bolzano–Weierstrass Theorem above, the Casorati–Weierstrass Theorem in *Unit 8*, and the Weierstrass M-test for uniform convergence.

2. (i) $(x+1)^2+y^2=1$. (ii) $5x-y+3=0$.

3. You have to tread with caution to avoid as much arithmetic as possible. One way is to write

$$|z-1| = 3-|z-2|$$

and so

$$|z-1|^2 = 9-6|z-2|+|z-2|^2$$

giving

$$36|z-2|^2 = (9+|z-2|^2-|z-1|^2)^2.$$

The right-hand side simplifies considerably and (after a few days' work) you should arrive at

$$8(x-\tfrac{3}{2})^2+9y^2=18.$$

(It is possible to go straight to x and y in the first equation, but in initial manipulations it is often simpler to work in terms of z.)

4. (i)

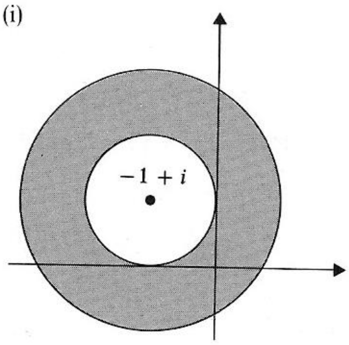


Fig. 34

(ii)

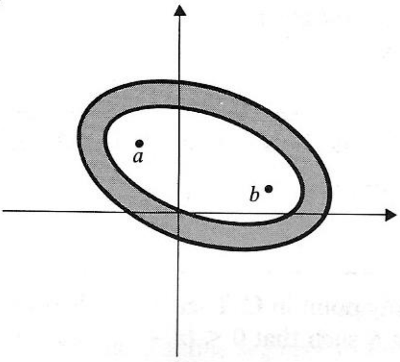


Fig. 35

(iii)

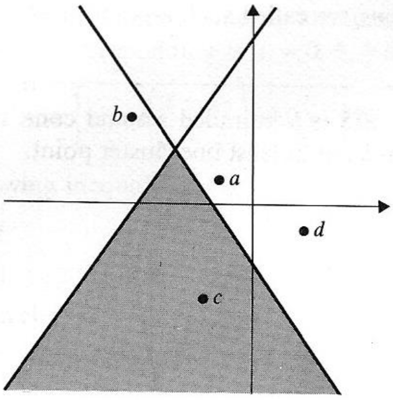


Fig. 36

$\left(\text{Note that } \left\{z : \left|\dfrac{z-a}{z-b}\right| \leqslant 1\right\}\right.$ is the set of all points whose distance from a is less than or equal to their distance from b.)

5. (i) The set is the set of points on the line segment joining z_1 and z_2, which is denoted by $[z_1, z_2]$ by analogy with an interval on the real line. The points corresponding to $t = 0, \frac{1}{3}, \frac{1}{2}, 1$ correspond to the points: z_2, the point of trisection of $[z_1, z_2]$ nearer to z_2, the midpoint, and z_1, respectively.

(ii) Since $\operatorname{Im}\left(\dfrac{z-z_2}{z_1-z_2}\right) = 0$, $t = \dfrac{z-z_2}{z_1-z_2}$ is real and $z = tz_1+(1-t)z_2$. Hence A is the line which includes the line segment of part (i).

6. (i) (a) $|\text{Re}\, z| = |x| \leqslant \sqrt{x^2+y^2}$ (because x and y are real)
$= |z|$.

(b) $|\text{Im}\, z| = |y| \leqslant \sqrt{x^2+y^2}$ (because x and y are real)
$= |z|$.

(c) Since, for any real numbers x and y, we have

$$(|x|-|y|)^2 \geqslant 0,$$

we can deduce that

$$|x|^2 + |y|^2 \geqslant 2|x| \cdot |y|,$$

and so the result follows.

(d) We have

$$\begin{aligned}(|\operatorname{Re} z| + |\operatorname{Im} z|)^2 &= (|x| + |y|)^2\\ &= |x|^2 + |y|^2 + 2|x| \cdot |y|\\ &\leqslant |x|^2 + |y|^2 + |x|^2 + |y|^2 \quad \text{(from (c))}\\ &= 2|z|^2.\end{aligned}$$

Also

$$|x|^2 + |y|^2 + 2|x| \cdot |y| \geqslant |x|^2 + |y|^2 = |z|^2.$$

So

$$(|\operatorname{Re} z| + |\operatorname{Im} z|)^2 \geqslant |z|^2.$$

The result follows because moduli are nonnegative.

(ii) For any complex number α,

$$\alpha + \bar{\alpha} = 2 \operatorname{Re} \alpha,$$

(see Problem 3(i) of Section 1.4).

Put $\alpha = z\bar{w}$ and note that $\overline{(\bar{w})} = w$; then

$$z\bar{w} + \bar{z}w = 2 \operatorname{Re}(z\bar{w}).$$

(iii) We have

$$\begin{aligned}(z + w)(\bar{z} + \bar{w}) &= z\bar{z} + w\bar{w} + z\bar{w} + \bar{z}w\\ &= |z|^2 + |w|^2 + z\bar{w} + \bar{z}w\\ &= |z|^2 + |w|^2 + 2\operatorname{Re}(z\bar{w}) \quad \text{(by (ii))}\\ &\leqslant |z|^2 + |w|^2 + 2|z\bar{w}| \quad \text{(by (i)(a))}\\ &= |z|^2 + |w|^2 + 2|zw|\\ &= |z|^2 + |w|^2 + 2|z| \cdot |w|\\ &= (|z| + |w|)^2.\end{aligned}$$

Also

$$(z + w)(\bar{z} + \bar{w}) = (z + w)\overline{(z + w)} = |z + w|^2.$$

Thus

$$|z + w|^2 \leqslant (|z| + |w|)^2, \quad \text{and so}$$
$$|z + w| \leqslant |z| + |w|,$$

since both sides are nonnegative.

(iv) From (iii), we have $|z| = |w + (z - w)| \leqslant |w| + |z - w|$. Thus $|z| - |w| \leqslant |z - w|$. Interchanging z and w, gives $|w| - |z| \leqslant |w - z|$ and combining the two results gives

$$\big||z| - |w|\big| \leqslant |z - w|.$$

The result corresponds to the fact that the difference between the lengths of any two sides of a triangle does not exceed the length of the third side.

7. (i) $d(z, w) = |z - w| = |w - z| = d(w, z)$.

(ii) If $z = x + iy$ and $w = u + iv$ then

$$d(z, w) = |z - w| = \sqrt{(x - u)^2 + (y - v)^2}.$$

Thus $d(z, w) = 0$ if and only if $x = u$ and $y = v$; that is if $z = w$.
$d(z, w) > 0$ otherwise, because $|z - w|$ is nonnegative.

(iii) We have

$$d(z, w) = |z - w| = |z - a + a - w|$$
$$\leqslant |z - a| + |a - w|$$

using the result of Problem 6(iii), with z replaced by $z - a$ and w replaced by $a - w$. Thus $d(z, w) \leqslant d(z, a) + d(a, w)$.

(Note that the set $\mathbf{C}$ and the function d with domain $\mathbf{C} \times \mathbf{C}$ constitutes a *metric space*.)

8. (i) $\frac{1}{2}(z\bar{w} + \bar{z}w) = \text{Re}(zw) \leqslant |zw|$ by part (a) of Problem 6(i).

(ii) Since

$$|z - w|^2 = (z - w)(\overline{z - w}) = (z - w)(\bar{z} - \bar{w}),$$

we know that

$$(z - w)(\bar{z} - \bar{w}) \geqslant 0$$

and so

$$z\bar{z} + w\bar{w} - w\bar{z} - z\bar{w} \geqslant 0,$$

from which

$$|z|^2 + |w|^2 \geqslant w\bar{z} + z\bar{w}.$$

9. (i) We know that $|z| = 2$ and we want to replace $\left|\frac{1}{z^2 + 1}\right| = \frac{1}{|z^2 + 1|}$ by something larger. This means replacing $|z^2 + 1|$ by something smaller. We can write $|z^2 + 1| = |z^2 - (-1)|$ and use the result of Problem 6(iv) to give $|z^2| - 1 \leqslant |z^2 + 1|$. So

$$\frac{1}{|z^2 + 1|} \leqslant \frac{1}{|z^2| - 1}$$
$$= \tfrac{1}{3}, \quad \text{since } |z| = 2.$$

(ii) A similar argument to that in (i) gives

$$\left|\frac{1}{3 + 4z}\right| \leqslant \frac{1}{4|z| - 3} = 1, \quad \text{since } |z| = 1.$$

10. Since S is bounded there is a closed disc with centre the origin and radius K, say, which contains S. Therefore there is a closed square, S_0, with vertices at $\pm K \pm iK$ which encloses S. The imaginary axis divides this square into two closed rectangles. At least one of these must contain infinitely many points. Choose this rectangle. If both rectangles contain infinitely many points, choose the right-hand one. The real axis divides this rectangle into two closed squares. At least one of these contains infinitely many points of S. Choose this one, or the upper one if they both do. Call this square S_1. If we continue in this way, dividing each square by a vertical line and then by a horizontal line, we get a nested sequence of closed squares $\{S_n\}$ where S_n has side of length $\frac{K}{2^n}$.

Also $\lim_{n\to\infty} \frac{K}{2^n} = 0$. By the Nested Rectangles Theorem, there is one and only one point z_0 in every square. This is a cluster point, because given $\varepsilon > 0$, the disc $|z - z_0| < \varepsilon$ contains the squares S_n where $\sqrt{2} \times \frac{K}{2^n} < \varepsilon$, that is, $2^n > \frac{K\sqrt{2}}{\varepsilon}$ (Fig. 37). (Notice that z_0 could be at the corner of S_n: this position giving the maximum size of disc required.) Thus the disc contains an infinite number of points of the set.

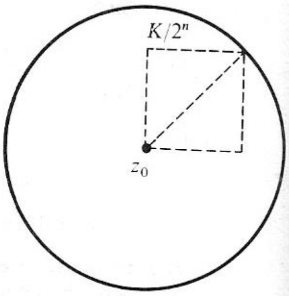


Fig. 37

Notice that there may be more than one cluster point. The cluster point will be unique if at each stage of the process only one of the two rectangles contains an infinite number of points of S.

Unit 2 Continuous Functions

Conventions

Before working through this text make sure you have read *A Guide to the Course: Complex Analysis.*

References to units of other Open University courses in mathematics take the form:

Unit M100 13, Integration II.

The set book for the course M231, Analysis, is M. Spivak, *Calculus*, paperback edition (W. A. Benjamin/Addison-Wesley, 1973). This is referred to as:

Spivak.

2.0 INTRODUCTION

In this unit we shall concentrate on continuous functions, both how to define them and the properties they have. Several of the definitions and techniques of real analysis generalize in a rather easy way: for example, the definition of the sum of functions and the techniques of proving results about limits. However, we shall concentrate less on such features of continuous functions, and more on properties which either have no genuine analogue in real analysis or are dealt with rather briefly in elementary courses in real analysis.

In particular, we shall spend some time on the question: "What does the domain of a reasonable function look like?". In complex analysis we can have domains of more complicated "shape" than the familiar intervals of real analysis. The above question in fact comes in two forms: one local, one global. The local version asks: "Where in the domain of a function is it reasonable to ask whether or not it has a limit?". Our answer is given in Sections 2.3 to 2.6. The global version asks: "What kinds of set must we replace intervals by to get analogues of the Intermediate Value Theorem and the Boundedness Theorem (see Chapter 7 of **Spivak**) for functions continuous on an interval?". We discuss this in Sections 2.7 to 2.10. All this work is preceded by some preliminaries on functions (in Sections 2.1 and 2.2).

You should expect and plan to spend a lot of your time on Sections 2.7 to 2.10. You may find the proofs of Theorems 7 and 11 hard; these may be omitted on a first reading, but you should aim to understand them in the long run.

Television and Radio

In the first television programme associated with the course we shall summarise several of the new definitions in this unit and illustrate the proof of Theorem 12.

In the first radio programme we shall give practical help with some of the more difficult manipulations of complex numbers and with the proof of Theorem 10.

2.1 FUNCTIONS

This section splits into two parts. In the first, which consists of a series of short notes, we discuss basic notions connected with the idea of function. In the second we discuss pictorial representations of functions, mainly in a series of worked examples. This topic will be taken up again later in the course.

* * * * * * * *

Functions

Since this course is about complex numbers, we shall on the whole use the term "function" to mean "function sending complex numbers to complex numbers". (In some contexts we use the term "complex function" instead.) Here is the precise definition.

Definition

A **function** is a set S of ordered pairs of complex numbers such that

$$\text{if } (z, w) \in S \text{ and } (z, w') \in S \text{ then } w = w'.$$

Notation for Functions

We usually denote functions by lower case letters such as f. Instead of $(z, w) \in f$, we normally write $f(z) = w$, and call w the **value of f at z**.

In practice functions are usually defined by formulas or by arrow notation, with appropriate restrictions on the sets involved. The following formulas define functions:

$f(z) = z + 2;$

$f(z) = 1/z, \quad z \neq 0;$

$f(z) = i \sin z, \quad z \in \mathbf{R}$ (recall that we regard $\mathbf{R}$ as subset of $\mathbf{C}$);

$f(z) = |z|^2, \quad |z| \leqslant 1.$

In arrow notation these would be:

$z \longrightarrow z + 2;$

$z \longrightarrow 1/z, \quad z \neq 0;$

$z \longrightarrow i \sin z, \quad z \in \mathbf{R};$

$z \longrightarrow |z|^2, \quad |z| \leqslant 1.$

Definition

The **domain** of f is the set of points to which f applies; that is

$$\{z \in \mathbf{C} : w = f(z) \text{ for some } w \in \mathbf{C}\}.$$

The **range** of f is the set of values that f takes, that is

$$\{w \in \mathbf{C} : w = f(z) \text{ for some } z \in \mathbf{C}\}.$$

If f has domain A and range contained in B, we can summarise this by writing $f: A \longrightarrow B$.

Finding the Domain

Initially we shall adopt the principle that **Spivak** follows: the domain is determined by the definition of the function. Thus the domains of the four functions above are, respectively,

$$\mathbf{C}, \quad \{z \in \mathbf{C} : z \neq 0\}, \quad \mathbf{R} \quad \text{and} \quad \{z \in \mathbf{C} : |z| \leqslant 1\}.$$

Finding the Range

This is not such an automatic process. It is not helpful to say that the range is

$$\{f(z) : z \text{ in the domain of } f\}.$$

We often have to find an *explicit* specification of this set, that is, put it in the form

$$\{z \in \mathbf{C} : P(z)\},$$

where $P(z)$ is some nice formula. The ranges of the four functions above are, respectively,

$$\mathbf{C},$$

$$\{z \in \mathbf{C} : z \neq 0\},$$

$$\{z \in \mathbf{C} : -1 \leqslant \operatorname{Im} z \leqslant 1, \operatorname{Re} z = 0\},$$

and the interval $[0, 1]$ of **R**.

Operations on Functions

Addition, $f + g$:

$$(f + g)(z) = f(z) + g(z).$$

Multiplication, $f \cdot g$:

$$(f \cdot g)(z) = f(z) \cdot g(z).$$

Division, f/g:

$$(f/g)(z) = f(z)/g(z).$$

Composition, $f \circ g$:

$$(f \circ g)(z) = f\big(g(z)\big).$$

It is easy to work out the domains of the combinations, given those of f and g. For example, if f has domain A and g has domain B, then $f + g$ and $f \cdot g$ have domain $A \cap B$.

Polynomial Functions

Starting from the *identity function* $z \longrightarrow z$ and the collection of *constant functions* $z \longrightarrow c$, for $c \in \mathbf{C}$, we can, using addition and multiplication, build up the class of polynomial functions. As you would expect from *Unit 1, Complex Numbers,* since $f(z) = a_0 + a_1 z + \ldots + a_n z^n$ is a polynomial in z of degree n, (for some non-negative integer n and complex numbers $a_0, \ldots, a_n$) f is called a **polynomial function**.

Rational Functions

We call f a **rational function** if there are polynomial functions p and q, q not the zero function, such that

$$f(z) = p(z)/q(z).$$

If q has degree not less than 1 then $q(\alpha) = 0$ for some α, by the Fundamental Theorem of Algebra, and so the domain of f does not include any such α.

Some Types of Function

If the range of f is contained in **R**, we call f *real-valued.*

If the domain of f is contained in **R**, we call f a *function of a real variable.*

If f is a real-valued function of a real variable, we call f a *real function.* (Real functions are the functions you deal with in a real analysis course.)

Standard Functions

$z \longrightarrow |z|$, the modulus function;

$\mathrm{Re}: z \longrightarrow \mathrm{Re}\, z$;

$\mathrm{Im}: z \longrightarrow \mathrm{Im}\, z$;

$z \longrightarrow \bar{z}$, the conjugation function.

The first three of these are real-valued.

The following compositions occur so frequently that it is common practice to abbreviate:

$(z \longrightarrow |z|) \circ f$ to $|f|$;

$\mathrm{Re} \circ f$ to $\mathrm{Re} f$;

$\mathrm{Im} \circ f$ to $\mathrm{Im} f$;

$(z \longrightarrow \bar{z}) \circ f$ to $\bar{f}$.

Real and Imaginary Parts

Since $f(z) = \mathrm{Re}(f(z)) + i\,\mathrm{Im}(F(z))$ for all z, we have

$$f = \mathrm{Re} f + i\,\mathrm{Im} f.$$

This shows that any function f can be decomposed into two real-valued functions: we often write U for $\mathrm{Re} f$ and V for $\mathrm{Im} f$. Since **C** and **R** × **R** are the same sets, we can regard U and V as having domains contained in **R** × **R**. In this context we often write $U(x + iy)$ as $U(x, y)$, and $V(x + iy)$ as $V(x, y)$.

For example, if $z = x + iy$, then

$$\begin{aligned} z^3 = (x + iy)^3 &= x^3 + 3x^2 \cdot iy - 3x \cdot y^2 - iy^3 \\ &= x(x^2 - 3y^2) + iy(3x^2 - y^2). \end{aligned}$$

Thus the function f given by $f(z) = z^3$ can be written as $f = U + iV$ where $U(x, y) = x(x^2 - 3y^2)$, $V(x, y) = y(3x^2 - y^2)$.

Conversely, given $U(x, y)$ and $V(x, y)$ as formulas in x and y, we can express $f(z)$ as a formula in z by using the equations $x = \frac{1}{2}(z + \bar{z})$, $y = -\frac{1}{2}i(z - \bar{z})$.

Pictorial Representation of Functions

In real analysis the graph of a function provides a very useful visualisation of its properties. However, the graph of a function $f: \mathbf{C} \longrightarrow \mathbf{C}$ is a subset of **C** × **C** (which is just **R** × **R** × **R** × **R**), and so a four-dimensional object. Thus we cannot expect to represent it very well on a two-dimensional page.

Nevertheless, it is often possible to get a good idea of the behaviour of f by the following approach. Draw two complex planes side by side. Traditionally we call the left-hand one the *z-plane* and the right-hand one the *w-plane*, because we normally write z for a typical point in the domain and w for the value of f at z. If we take a point z_0 in the z-plane, we can locate the value $w_0 = f(z_0)$ in the

w-plane. We could do this for a whole set S of points in the z-plane and look at the corresponding set of values in the w-plane. It is useful to have a name for this set.

Definition

> Let f be a function, and S a subset of the domain of f. Then the **image of S under** f, denoted by $f(S)$, is the set
>
> $$\{f(z) : z \in S\}.$$

Note that if S *is* the domain then $f(S)$ is the range.

On intuitive grounds you would expect that if S is a curve and z moves along S, then $f(z)$ will trace out a curve in the w-plane. It is usual to take easily visualized curves in the z-plane as the set S; for example, lines parallel to one or other axis.

Example 1

Let $f(z) = z^2$. Find the image under f of the line $S_a = \{(a, y) : y \in \mathbf{R}\}$.

Solution

Since $f(x, y) = (x + iy)^2 = (x^2 - y^2) + 2ixy$, we have $f = U + iV$ where

$$U(x, y) = x^2 - y^2, \quad V(x, y) = 2xy.$$

Let (u, v) be a typical point in the w-plane. Then

$$\begin{aligned}(u, v) \in f(S_a) &\Leftrightarrow u = U(x, y) \text{ and } v = V(x, y) \text{ for some } (x, y) \in S_a \\ &\Leftrightarrow u = a^2 - y^2 \text{ and } v = 2ay \text{ for some } y \in \mathbf{R}.\end{aligned}$$

We have now to distinguish two cases.

(i) $a = 0$: then

$$(u, v) \in f(S_a) \Leftrightarrow u = -y^2 \text{ and } v = 0 \text{ for some } y \in \mathbf{R}.$$

Hence $f(S_a)$ is the non-positive real axis.

(ii) $a \neq 0$: then

$$(u, v) \in f(S_a) \Leftrightarrow u = a^2 - \left(\frac{v}{2a}\right)^2.$$

Hence $f(S_a)$ is the parabola $u = a^2 - v^2/(4a^2)$ in the w-plane.

Once we have the images of lines, we can find the images of strips by building them out of lines.

Example 2

Sketch the image of the strip $S = \{(x, y) : 1 < x < 2, y \in \mathbf{R}\}$ under the function $f(z) = z^2$.

Solution

$$\begin{aligned}(u, v) \in f(S) &\Leftrightarrow (u, v) = f\big((a, y)\big) \text{ for some } a \in (1, 2) \text{ and } y \in \mathbf{R} \\ &\Leftrightarrow (u, v) \in f(S_a) \text{ for some } a \in (1, 2) \\ &\Leftrightarrow u = a^2 - \frac{v^2}{4a^2} \text{ for some } a \in (1, 2).\end{aligned}$$

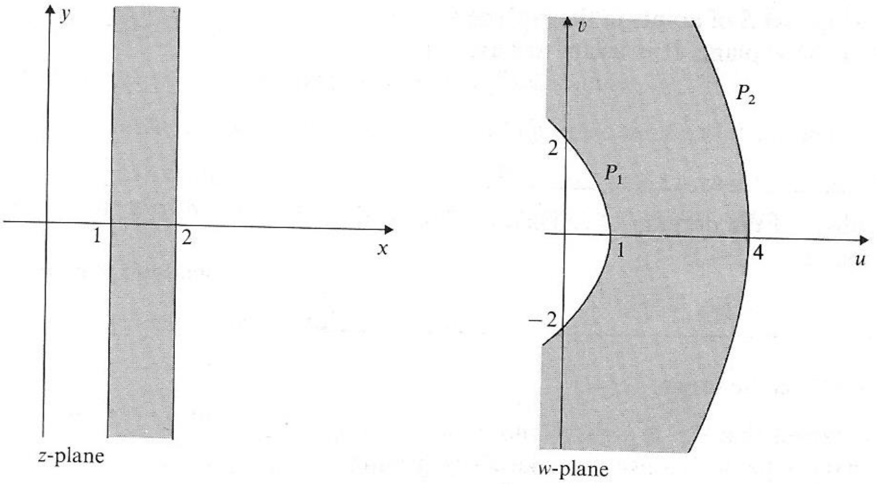


Fig. 1

Thus $f(S)$ is the set consisting of the parabolas $P_a = \left\{(u, v) : u = a^2 - \dfrac{v^2}{4a^2}\right\}$ for $a \in (1, 2)$. Now if $1 < a < 2$ then P_a lies between the parabolas P_1 and P_2. Conversely, it can be shown that any point between P_1 and P_2 lies on some P_a. Thus $f(S)$ is the set indicated on the w-plane in Fig. 1.

As you can see from the examples it is rather difficult to calculate the images of sets under functions. A somewhat easier but useful procedure is to calculate inverse images, which we define as follows.

Definition

If S is any subset of $\mathbf{C}$ and f is a function, then the **inverse image of S under f** is the set $\{z : f(z) \in S\}$. This set is denoted by $f^{-1}(S)$; thus

$$f^{-1}(S) = \{z : f(z) \in S\}.$$

In other words, $z \in f^{-1}(S) \Leftrightarrow f(z) \in S$.

Example 3

Find the inverse image under the function $f(z) = z^2$ of the line $L = \{(1, v) : v \in \mathbf{R}\}$.

Solution

$$\begin{aligned}(x, y) \in f^{-1}(L) &\Leftrightarrow x^2 - y^2 = 1 \text{ and } 2xy = v \text{ for some } v \in \mathbf{R}\\ &\Leftrightarrow x^2 - y^2 = 1.\end{aligned}$$

Thus $f^{-1}(L)$ is the hyperbola $x^2 - y^2 = 1$ in the z-plane (Fig. 2).

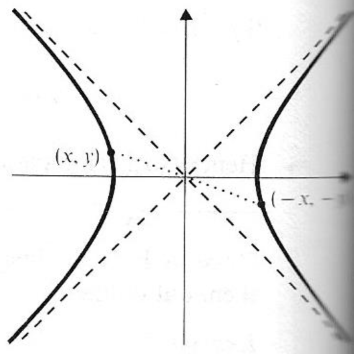


Fig. 2

Example 4

Find the inverse image under the function $f(z) = z^2$ of the disc $D = \{w : |w| < 4\}$.

Solution

$$\begin{aligned}z \in f^{-1}(D) &\Leftrightarrow f(z) \in D\\ &\Leftrightarrow |z^2| < 4\\ &\Leftrightarrow |z|^2 < 4\\ &\Leftrightarrow |z| < 2.\end{aligned}$$

Thus $f^{-1}(D) = \{z : |z| < 2\}$.

In many cases we can actually calculate images by using inverse images, as follows. Recall that a function is called **one-one** if whenever a, b are in the domain of f and $f(a) = f(b)$ then $a = b$. Thus f is one-one precisely if the inverse image of each point in its range consists of just one point. The inverse* of f, denoted by f^{-1}, is the set $f^{-1} = \{(w, z) : w = f(z)\}$. If f is one-one then $g = f^{-1}$ is a *function* and the image $g(S)$ of a set S under g is just the inverse image $f^{-1}(S)$ of S under f. Hence if f is one-one (and therefore f^{-1} is a function), we can view the notation $f^{-1}(S)$ as meaning either

(i) the image of S under f^{-1}

or

(ii) the inverse image of S under f.

Note that $f^{-1}(S)$ is defined even when f^{-1} is not a function.

Example 5

Find f^{-1} when $f(z) = \dfrac{1}{1 + z}$.

Solution

$$\begin{aligned} w = \frac{1}{1+z} &\Leftrightarrow (1 + z)w = 1 \text{ and } z \neq -1 \\ &\Leftrightarrow 1 - w = wz \text{ and } z \neq -1 \\ &\Leftrightarrow z = \frac{1 - w}{w} \text{ and } w \neq 0. \end{aligned}$$

Hence, $f^{-1}(w) = \dfrac{1 - w}{w}$, that is, $f^{-1}(z) = \dfrac{1 - z}{z}$.

Example 6

Find the image of the disc $D = \{z : |z| \leqslant \frac{1}{2}\}$ under the function $f(z) = \dfrac{1}{1 + z}$.

Solution

The inverse g of f is $g(z) = \dfrac{1 - z}{z}$. Thus

$$\begin{aligned} w \in f(D) &\Leftrightarrow w \in g^{-1}(D) \\ &\Leftrightarrow g(w) \in D \\ &\Leftrightarrow \left|\frac{1 - w}{w}\right| \leqslant \tfrac{1}{2} \\ &\Leftrightarrow |w - 1| \leqslant \tfrac{1}{2}|w|. \end{aligned}$$

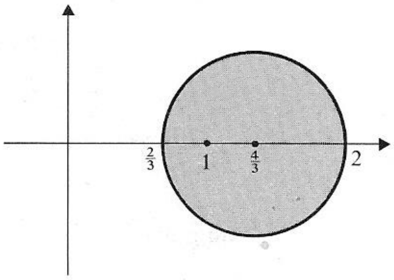


Fig. 3

The set $\{w : |w - 1| \leqslant \frac{1}{2}|w|\}$ can be shown to be the closed disc illustrated in Fig. 3, but we shall omit the details.

---

* In the Foundation Course the inverse of f is called the reverse mapping of f.

Summary

In this section we have discussed the notion of function and associated concepts, and introduced the topic of pictorial representation.

Self-Assessment Questions

1. Find the domains of the following functions.

   (i) $f(z) = \dfrac{1}{|z| - 3}$.

   (ii) $f(z) = \sqrt{\operatorname{Re} z - 2}$ (where $\sqrt{\ }$ is the square root of real analysis).

   (iii) $f(z) = \dfrac{3z}{(z-1)(z-2)}$.

   (iv) $f(z) = \dfrac{z}{1/z}$.

2. If $f(z) = \dfrac{z}{z+4}$ and $z = i$, evaluate

   (i) $\bar{f}(\bar{z} - 1)$;

   (ii) $|f|(z + 1)$;

   (iii) $\dfrac{1}{\bar{f}}(\bar{z})$.

3. Express the function $f(z) = 1/z$ in the form

$$f(x + iy) = U(x, y) + iV(x, y).$$

4. Is $z \longrightarrow \operatorname{Re} z + \operatorname{Im} z$ one-one?

5. What is the inverse image of $L = \{(u, v) : u \in \mathbf{R}, v = 2\}$ under the function $f(z) = z^2$?

6. A function f with domain $\mathbf{C}$ is said to be *odd* if $f(z) = -f(-z)$ for all z, and *even* if $f(z) = f(-z)$ for all z.

   Classify the following functions as odd, even, or neither odd nor even.

   (i) $f : z \longrightarrow |z|$.

   (ii) $f : z \longrightarrow \operatorname{Re} z$.

   (iii) $f : z \longrightarrow \operatorname{Im} z$.

   (iv) $f : z \longrightarrow \bar{z}$.

Solutions

1. (i) $\{z : |z| \neq 3\}$. (ii) $\{z : \operatorname{Re} z \geqslant 2\}$. (iii) $\{z : z \neq 1, 2\}$. (iv) $\{z : z \neq 0\}$.

2. (i) $\dfrac{-1 + 2i}{5}$. (ii) $\dfrac{\sqrt{13}}{13}$. (iii) $1 - 4i$.

3. $\dfrac{1}{x + iy} = \dfrac{x - iy}{x^2 + y^2}$; thus $U(x, y) = \dfrac{x}{x^2 + y^2}$, $V(x, y) = \dfrac{-y}{x^2 + y^2}$.

4. No; for example, $\operatorname{Re} 0 + \operatorname{Im} 0 = 0$, and $\operatorname{Re}(1 - i) + \operatorname{Im}(1 - i) = 1 - 1 = 0$.

5. $(x, y) \in f^{-1}(L) \Leftrightarrow x^2 - y^2 = u$ and $2xy = 2$ for some $u \in \mathbf{R}$

$$\Leftrightarrow xy = 1.$$

   Thus $f^{-1}(L)$ is the hyperbola $xy = 1$ in the z-plane.

6. (i) Even. (ii) Odd. (iii) Odd. (iv) Odd.

2.2 PROBLEMS

1. Show that f is one-one, and calculate f^{-1}, when

$$f(z) = \frac{z+1}{3z-2}.$$

2. Find the images of the strip $A = \{z : 0 \leqslant \operatorname{Re} z \leqslant 1\}$ and the circle $C = \{z : |z| = 1\}$ under the functions

(i) $f(z) = (2+i)z + 3$,

(ii) $g(z) = z^2$.

We say that a function f with domain $\mathbf{C}$ is an **isometry** if $|f(z_1) - f(z_2)| = |z_1 - z_2|$ for all $z_1, z_2 \in \mathbf{C}$. (This concept will appear in a number of problems throughout the unit.)

---

3. (a) Show that the composition of two isometries is an isometry.

(b) Show that any function of the form $f(z) = az + b$ or $g(z) = a\bar{z} + b$, where $|a| = 1$, is an isometry.

(c) Show that $z \longrightarrow a\bar{z} + b$, where $|a| = 1$, is the composition of a reflection, rotation, and translation. (A function f with domain $\mathbf{C}$ is called a **reflection** (in the real axis) if $f(z) = \bar{z}$, a **rotation** (about 0) if $f(z) = az$, where a is a fixed complex number with $|a| = 1$, a **translation** if $f(z) = z + z_0$, where z_0 is a fixed complex number.)

In fact it can be shown that *any* isometry has one of the forms in Problem 3(b). This is worth remembering; we omit the *proof* since it is a little cumbersome.

Solutions

1. $$w = \frac{z+1}{3z-2} \Leftrightarrow (3z-2)w = z + 1$$

$$\Leftrightarrow (3w-1)z = 2w + 1$$

$$\Leftrightarrow z = \frac{2w+1}{3w-1}.$$

This shows that f is one-one, and that $f^{-1}(w) = \dfrac{2w+1}{3w-1}$.

2. (i) f is one-one, since $w = (2+i)z + 3 \Leftrightarrow z = \dfrac{w-3}{2+i}$.

Thus

$$0 \leqslant \operatorname{Re} z \leqslant 1 \Leftrightarrow 0 \leqslant \operatorname{Re}\left(\frac{w-3}{2+i}\right) \leqslant 1.$$

Let $w = u + iv$. Then

$$\frac{w-3}{2+i} = \frac{2-i}{5}[(u-3) + iv],$$

and so

$$\operatorname{Re}\left(\frac{w-3}{2+i}\right) = \frac{2}{5}(u-3) + \frac{1}{5}v.$$

Hence $0 \leqslant \operatorname{Re} z \leqslant 1 \Leftrightarrow 0 \leqslant 2(u-3) + v \leqslant 5$. In other words, $f(A)$ is the set of points between the lines given by

$$2u + v = 6 \text{ and } 2u + v = 11.$$

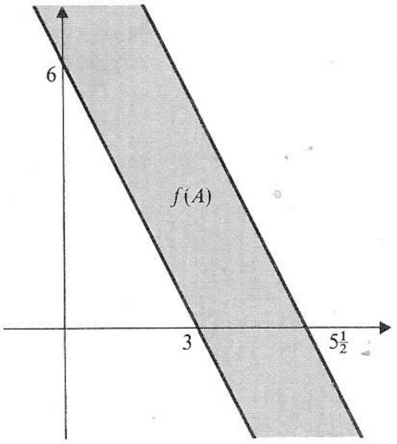


Fig. 4

Also $|z| = 1 \Leftrightarrow \left|\dfrac{w-3}{2+i}\right| = 1 \Leftrightarrow |w-3| = \sqrt{5}$. Thus $f(C)$ is the circle with centre 3 and radius $\sqrt{5}$.

(ii) We proceed much as in Example 2 of Section 2.1. $(u, v) \in f(A)$ if and only if (u, v) lies in the image of the line $\{(a, y) : y \in \mathbf{R}\}$ for some $a \in [0, 1]$.

If $a > 0$, then $u = a^2 - v^2/(4a^2)$.

If $a = 0$, then $u \leqslant 0$, $v = 0$ (from Example 1 of Section 2.1). Hence $f(A)$ is the set of points on and to the left of the parabola $u = 1 - v^2/4$.

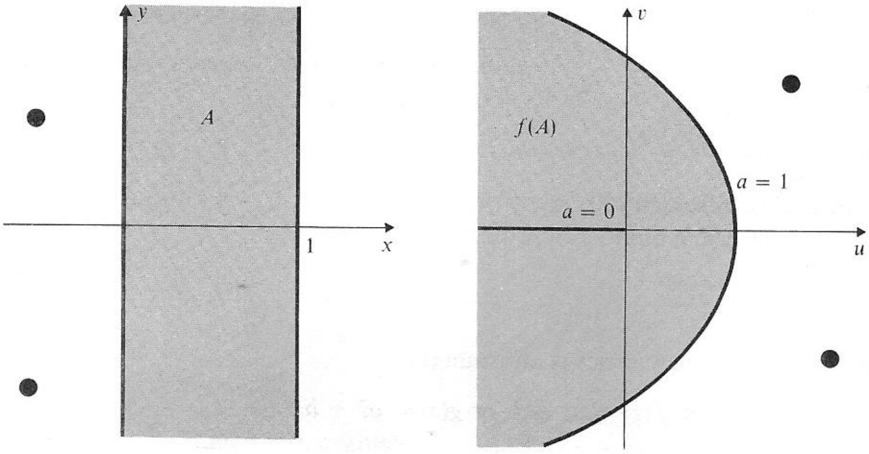


Fig. 5

Since $|z| = 1 \Leftrightarrow |z|^2 = 1$, $f(C)$ is just C.

3. (a) Let f and g be isometries. Then

$$\begin{aligned}|(f \circ g)(z_1) - (f \circ g)(z_2)| &= |f(g(z_1)) - f(g(z_2))| \\ &= |g(z_1) - g(z_2)| \quad \text{(because } f \text{ is an isometry)} \\ &= |z_1 - z_2| \quad \text{(because } g \text{ is an isometry).}\end{aligned}$$

Thus $f \circ g$ is an isometry.

(b)
$$\begin{aligned}|(az_1 + b) - (az_2 + b)| &= |a(z_1 - z_2)| \\ &= |a| \cdot |z_1 - z_2| \\ &= |z_1 - z_2| \quad \text{(because } |a| = 1\text{).}\end{aligned}$$

Thus $f(z) = az + b$ is an isometry.

Also
$$\begin{aligned}|(a\bar{z}_1 + b) - (a\bar{z}_2 + b)| &= |a| \cdot |\overline{z_1 - z_2}| \\ &= |z_1 - z_2| \quad \text{(because } |a| = 1\text{).}\end{aligned}$$

Thus $g(z) = a\bar{z} + b$ is an isometry.

(c) $z \longrightarrow a\bar{z} + b$, where $|a| = 1$, is the composition

$$(z \longrightarrow z + b) \circ (z \longrightarrow az) \circ (z \longrightarrow \bar{z}).$$

Now $z \longrightarrow \bar{z}$ is a reflection (in the x-axis), $z \longrightarrow az$, $|a| = 1$, a rotation (through Arg a), and $z \longrightarrow z + b$ a translation.

2.3 LIMITS, CONTINUITY AND OPEN SETS

The first part of this section discusses limits in the context of complex analysis, and is a fairly close analogue of the situation in real analysis. The second part discusses interior points and open sets, which you may not have studied in much detail previously, and so this material is the part to concentrate on.

* * * * * * * *

We hope that you recall the definition of "limit" in real analysis. We say that f has limit l near a if for every $\varepsilon > 0$ there is some $\delta > 0$ such that, for all x, if $0 < |x - a| < \delta$, then $|f(x) - l| < \varepsilon$. Think for a moment how you would change this to a complex analysis definition. Obviously you would preserve the logical form, and make at least some of the variables range over complex numbers. Since ε and δ measure smallness, they had better stay as they are, positive real numbers. Obviously a and l should be complex, as should x, and so we replace it by the traditional z. Conveniently, the modulus signs still apply (and agree with the old versions if the quantities inside them are real).

Definition

Let f be a function, and a and l be complex numbers. We say that f **has limit** l **near** a if for every $\varepsilon > 0$ there is some $\delta > 0$ such that, for all z,

if $0 < |z - a| < \delta$, then $|f(z) - l| < \varepsilon$.

(Note that it is common practice to say "at a" instead of "near a", but this could be misleading since we do not demand that $f(a)$ should be close to l nor even that f be defined at a.)

As in real analysis, we can prove that a function cannot have two distinct limits at a. (The proof goes through as for real analysis. See **Spivak**, Theorem 5-1.) This result justifies the following notation. If f has limit l near a, we write $\lim\limits_{z \to a} f(z)$ for l. This notation is unambiguous because l is unique.

The technique of proving results about limits is very similar to that in real analysis. A simple example is the result that $\lim\limits_{z \to a} f(z) = c$ if f is the constant function $z \longrightarrow c$. For, given any $\varepsilon > 0$, let $\delta = 1$: if $0 < |z - a| < \delta$ then

$$|f(z) - c| = |c - c| = 0.$$

A similar argument, with $\delta = \varepsilon$, shows that $\lim\limits_{z \to a} z = a$. Here is a more complicated example.

Example 1

Show that $\lim\limits_{z \to a} z^2 = a^2$.

Solution

By analogy with techniques of real analysis we try to find an inequality

$$|z^2 - a^2| \leqslant k|z - a|$$

for some positive k, and sufficiently small non-zero $|z - a|$. Now

$$\begin{aligned} |z^2 - a^2| &= |z + a| \cdot |z - a| \\ &= |(z - a) + 2a| \cdot |z - a| \\ &\leqslant (1 + 2|a|)|z - a| \quad \text{if } |z - a| < 1. \end{aligned}$$

Thus let $\delta = \min\left(1, \dfrac{\varepsilon}{1 + 2|a|}\right)$: if $|z - a| < \delta$ then $|z^2 - a^2| < \varepsilon$.

You could spend many happy hours proving similar but more complicated results. But fairly soon you probably would feel the need for some more systematic techniques for obtaining these. In fact, just as in real analysis, we can prove a general theorem which describes how the limit operator relates to the various operations on the complex plane.

Theorem 1

Suppose that $\lim_{z \to a} f(z) = l$ and $\lim_{z \to a} g(z) = m$. Then:

(i) $\lim_{z \to a} (f + g)(z) = l + m$;

(ii) $\lim_{z \to a} (f \cdot g)(z) = l \cdot m$;

(iii) $\lim_{z \to a} (f/g)(z) = l/m$, if $m \neq 0$;

(iv) $\lim_{z \to a} \bar{f}(z) = \bar{l}$;

(v) $\lim_{z \to a} |f|(z) = |l|$

(vi) $\lim_{z \to a} (\operatorname{Re} f)(z) = \operatorname{Re} l$;

(vii) $\lim_{z \to a} (\operatorname{Im} f)(z) = \operatorname{Im} l$.

Proof

The proofs of (i) and (ii) are virtually identical to the real analysis case (see Theorem 5-2 of **Spivak**). For (iv), it suffices to observe that $|\bar{f}(z) - \bar{l}| = |f(z) - l|$. Also, (v) holds since

$$||f|(z) - |l|| = ||f(z)| - |l|| \leqslant |f(z) - l|.$$

(See Problem 6 of Section 1.6 of *Unit 1*.)

For (vi),

$$\begin{aligned} \lim_{z \to a} (\operatorname{Re} f)(z) &= \tfrac{1}{2} \lim_{z \to a} (f + \bar{f})(z) \quad \text{by (ii)}, \\ &= \tfrac{1}{2}(l + \bar{l}) \quad \text{by (i) and (iv)}, \\ &= \operatorname{Re} l; \end{aligned}$$

similarly for (vii).

Case (iii) requires a little more thought, and careful application of the definition of limit. Note first that by (ii) it is enough to prove that $\lim_{z \to a} (1/g)(z) = 1/m$, if $m \neq 0$.

Now

$$\begin{aligned} \left|\frac{1}{g}(z) - \frac{1}{m}\right| &= \left|\frac{1}{g(z)} - \frac{1}{m}\right| \\ &= \frac{|g(z) - m|}{|m| \cdot |g(z)|}. \end{aligned}$$

We would like this last expression to be bounded above by $k|g(z) - m|$ for some k, and small enough non-zero $|z - a|$. This requires that $|g(z)|$ be bounded below by some sufficiently small non-zero number, for small enough non-zero $|z - a|$. We require this for another purpose: since we want $1/g(z)$ to be defined for small enough $|z - a|$, this requires that $g(z) \neq 0$.

Now (by taking $\varepsilon = \frac{1}{2}|m|$) there is $\delta_0 > 0$ such that whenever $0 < |z - a| < \delta_0$, we have

$$|g(z) - m| < \tfrac{1}{2}|m|, \text{ and so } |g(z)| > \tfrac{1}{2}|m| > 0;$$

thus

$$\frac{|g(z) - m|}{|m| \cdot |g(z)|} \leqslant \frac{2}{|m|^2} |g(z) - m|.$$

Thus we have found a k, namely $k = 2/|m|^2$. It is now easy to finish off the proof: given $\varepsilon > 0$ there is a $\delta_1 > 0$ such that

$$|g(z) - m| < \tfrac{1}{2}|m|^2\varepsilon \text{ whenever } 0 < |z - a| < \delta_1.$$

If $0 < |z - a| < \min(\delta_0, \delta_1)$ then

$$\left|\frac{1}{g}(z) - \frac{1}{m}\right| \leqslant \frac{2}{|m|^2}|g(z) - m| < \varepsilon. \quad \blacksquare$$

There is a corollary of Theorem 1 that is quite useful if one knows two-variable calculus, since it reduces limits of (complex) functions to limits of functions of two real variables.

Corollary

Let $f = U + iV$ (where U and V are real-valued), and $\lambda = l + im$, where l and m are real. Then $\lim_{z \to a} f(z) = \lambda$ if and only if

$$\lim_{z \to a} U(z) = l \text{ and } \lim_{z \to a} V(z) = m.$$

Proof

"Only if" follows from (vi) and (vii) of the theorem; "If" follows from (i) and a trivial use of (ii). $\blacksquare$

If you know two-variable calculus, note that $\lim_{z \to a} U(z)$ is just the double limit $\lim_{(x,y) \to (b,c)} U(x, y)$, where a is the point $(b, c) \in \mathbf{R} \times \mathbf{R}$. See the Appendix of *Unit M231 3, Limits and Continuity*.)

Definition

We say that f is **continuous at** a if $\lim_{z \to a} f(z) = f(a)$.

We sometimes say that f is *continuous* if f is continuous at each point a of its domain, but this can become confusing when the domain is complicated, so that we shall not use the phrase in *precise* contexts. See Section 2.5 for more on this.

Clearly each constant function $z \longrightarrow c$ and the identity function $z \longrightarrow z$ are continuous, and by applying parts (iv)–(vii) of Theorem 1 it follows immediately that the standard functions $z \longrightarrow \bar{z}$, $z \longrightarrow |z|$, Re and Im are continuous. As regards parts (i), (ii) and (iii) it is worth restating them in terms of continuity.

Theorem 2

Let f and g be continuous at a. Then:

(i) $f + g$ is continuous at a;

(ii) $f \cdot g$ is continuous at a;

(iii) f/g is continuous at a, provided that $g(a) \neq 0$.

It follows easily by induction on n that if $f_1, \ldots, f_n$ are continuous at a then $f_1 + \ldots + f_n$ and $f_1 \cdot \ldots \cdot f_n$ are also continuous at a. Hence any polynomial function is continuous. It then follows that any rational function is continuous.

You may have wondered why we did not write down in Theorem 2 analogues of clauses (iv)–(vii) of Theorem 1. This is because there is another way of combining continuous functions, which has no natural analogue for limits, is of wide application, and will imply these results. We shall show that the *composition* of continuous functions is continuous.

In order to prove this, it is worth reformulating the definition of limit as it applies to continuity.

If $\lim_{z \to a} f(z) = f(a)$, then for every $\varepsilon > 0$ there is $\delta > 0$ such that, for all z,

$$\text{if } 0 < |z - a| < \delta, \text{ then } |f(z) - f(a)| < \varepsilon.$$

But you see that the restriction that $z \neq a$, so necessary for the definition of *limit* (as is explained at length in **Spivak** for the real analysis case), is unnecessary for *continuity*; for certainly if $z = a$, then $|f(z) - f(a)| = 0 < \varepsilon$. Thus we have the following equivalence:

f is continuous at $a \Leftrightarrow$ for every $\varepsilon > 0$ there is $\delta > 0$ such that, for all z, if $|z - a| < \delta$ then $|f(z) - f(a)| < \varepsilon$.

Using this we can prove easily the following result on composition. (The proof is the precise analogue of the one in real analysis. See **Spivak**, Theorem 6-2.)

Theorem 3

If f is continuous at $g(a)$ and g is continuous at a, then $f \circ g$ is continuous at a.

From this theorem we can immediately deduce that if g is continuous at a then $\bar{g}$, $|g|$, $\operatorname{Re} g$ and $\operatorname{Im} g$ are all continuous at a (since, for example, $\operatorname{Re} g = \operatorname{Re} \circ g$ and Re is continuous at $g(a)$). Conversely, if $\operatorname{Re} g$ and $\operatorname{Im} g$ are continuous at a so is g (since $g = \operatorname{Re} g + i\operatorname{Im} g$).

So far we have not been very concerned about the *domains* of the functions we have been dealing with. Yet there is one technical aspect of the definition of limit which is highly relevant to this. It comes over most clearly when we consider continuity.

Suppose that f is continuous at a. Then for every $\varepsilon > 0$ there is some $\delta > 0$ such that

$$\text{for all } z \text{ in } \{\zeta : |\zeta - a| < \delta\}, \text{ we have } |f(z) - f(a)| < \varepsilon.$$

But in order for this last line to make sense, we must have $\{\zeta : |\zeta - a| < \delta\}$ contained in the domain of f. Hence in order for the *whole* definition to make sense, we must have $\{\zeta : |\zeta - a| < \delta_0\}$ contained in the domain of f, for some particular $\delta_0 > 0$. (Note that any smaller δ_0 would do also.)

Definition

Let S be a set, and a be a point of S. We say that a is an **interior point of** S if $\{\zeta : |\zeta - a| < \delta_0\} \subseteq S$ for some $\delta_0 > 0$.

Thus, when we say that f is continuous at a point a, there is an implicit requirement that a is an interior point of the domain of f. Hence very often the δ in the definition of limits or continuity comes in two parts. You saw this in the proof of Theorem 1, part (iii), where we took δ_0 to ensure that $\frac{1}{g}$ was defined on $\{\zeta : 0 < |\zeta - a| < \delta_0\}$ and δ_1 to make $|g(z) - m|$ small, and finally δ to be the minimum of δ_0 and δ_1.

In fact the definition of "f has limit l near a" does not require f to be even defined *at* a, and so we need only require that a is an interior point of $S \cup \{a\}$, where S is the domain of f. In this case a is said to be a **punctured interior point of** S, but we shall not often require this concept: interior points are much more important. The set $\{z : 0 < |z - z_0| < \delta\}$ is called a **punctured disc** centred at z_0.

Note that if a is an interior point of S and of T, then a is an interior point of $S \cap T$; for if $z \in S$ whenever $|z - a| < \delta_0$ and $z \in T$ whenever $|z - a| < \delta_1$ then $z \in S \cap T$ whenever $|z - a| < \min(\delta_0, \delta_1)$.

A similar statement holds for punctured interior points. Now if the domain of f is S and the domain of g is T, then $S \cap T$ is the domain of $f + g$, and of $f \cdot g$. Thus clauses (i) and (ii) of Theorems 1 and 2 remain valid even when we make explicit this assumption that a is a punctured interior point. Of course in clause (iii) we actually proved it.

Because of this occurrence of interior points in the definition of limit, *and* because of our desire to extend the definition of limit to points other than interior points of the domain, we now introduce some *topological* ideas.

Definition

> Let S be any set (in $\mathbf{C}$). Then S is called **open** if every point of S is an interior point of S.

In other words, S is open precisely if for each $a \in S$ there is $\delta > 0$ such that the disc $\{z : |z - a| < \delta\}$ is contained in S. Incidentally, we shall call that disc *the δ-neighbourhood of a*, or just *a neighbourhood of a*.

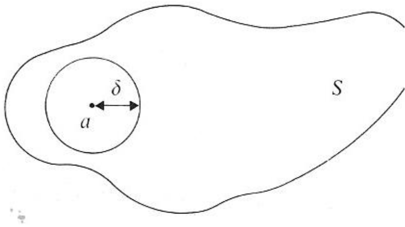


Fig. 6

Note that δ can vary with a: in particular δ will have to be small near the "edge" of S. But if S is unbounded and so has no edge, δ is not constrained to be small. For example, the set $\mathbf{C}$ is open—given any $a \in \mathbf{C}$, *any* δ-neighbourhood of a lies in $\mathbf{C}$. Also, by use of those curious "vacuous" arguments, the empty set $\varnothing$ is open. (Find a point in $\varnothing$ which is *not* an interior point!)

Apart from these extreme cases, there are many more open sets. In fact, we shall now show that all the sets which were called "open" in *Unit 1*, such as open disc, open rectangle, open annulus, etc., are in fact open in this *precise* sense. Instead of attacking them indiscriminately, it is best (as always) to be systematic, beginning with simple open sets and then proving some theorems about combining open sets.

Example 2

Show that the following sets are open:

(i) any open disc;

(ii) the set $\{z : |z - b| > r\}$ where $r \geqslant 0$—that is, the outside of a circle for $r \neq 0$, and the set $\{z : z \neq b\}$ for $r = 0$;

(iii) the open half-plane $\{(x, y) : y > 0\}$.

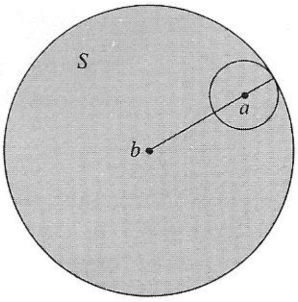


Fig. 7

Solution

(i) Let $S = \{z : |z - b| < r\}$ (where $r > 0$), and $a \in S$ (Fig. 7).

Clearly $|a - b| < r$ and so, if we put $\delta = r - |a - b|$, then $\delta > 0$. If $|z - a| < \delta$, then

$$|z - b| \leqslant |z - a| + |a - b| < \delta + (r - \delta) = r.$$

(This is just the analytic translation of obvious geometric facts.)

(ii) The above solution is a bit over-thorough for general use, and we shall be a little briefer from now on. Let $S = \{z : |z - b| > r\}$ ($r > 0$), $a \in S$ (Fig. 8). If $\delta = |a - b| - r$, then $\{z : |z - a| < \delta\} \subseteq S$.

This covers the case $r = 0$ also.

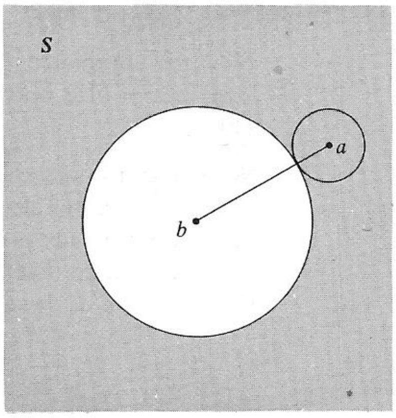


Fig. 8

(iii) Let $S = \{(x, y): y > 0\}$. If $a \in S$, then $\{z: |z - a| < \operatorname{Im} a\} \subseteq S$.

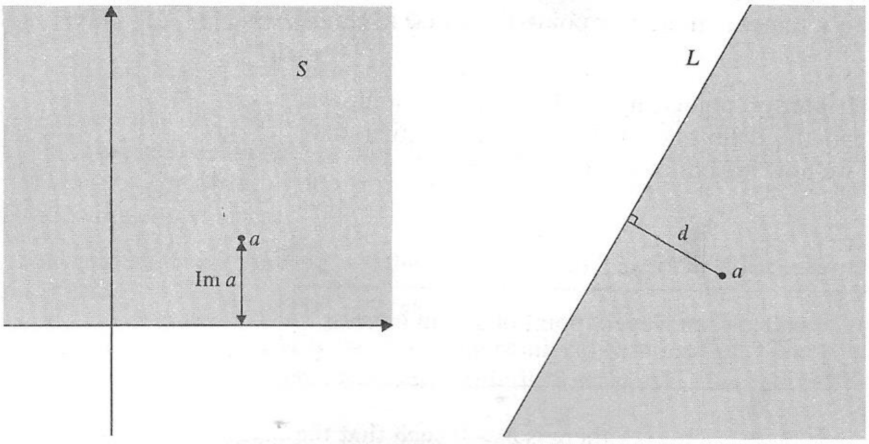


Fig. 9

For a *general half-plane*, consisting of all points to one side of a line L (Fig. 9), we must replace $\operatorname{Im} a$ by the distance d of a from L. (See also Problem 5(b) of Section 2.6 for another way of deriving the general case.)

We can prove in the same manner that many other sets are open, but the details get more complicated as the specification of the sets does: so we shall consider just one more example of this sort.

Example 3

Show that the open rectangle $R = \{z: a < \operatorname{Re} z < b, c < \operatorname{Im} z < d\}$ is in fact open.

Solution

Let $w = u + iv \in R$; then if δ is the smallest of the values $u - a$, $b - u$, $v - c$, $d - v$, it is obvious that $\{z: |z - w| < \delta\} \subseteq R$.

As we have said, there is a more systematic way of obtaining results such as Example 3, by considering the operations which preserve open sets. There are just three operations on sets that one normally considers:

intersection $A \cap B = \{z: z \in A \text{ and } z \in B\}$,

union $A \cup B = \{z: z \in A \text{ or } z \in B\}$, and

complement $\mathbf{C} - A = \{z \in \mathbf{C}: z \notin A\}$.

Now the complement of an open set is not necessarily open; for, by Example 2(ii), $\{z: z \neq b\}$ is open, but it is easy to see that $\{z: z = b\}$ (the single point b) is not.

What about intersection? Let us look back at Example 3 and try to generalize it. An open rectangle is the intersection of four half-planes (Fig. 10). For each half-plane we took a δ which made (u, v) an interior point of it: for example, $\delta = u - a$ ensured that (u, v) was an interior point of $\{(x, y): x > a\}$. Then we took the minimum of these four δ's. This is exactly the method we use in general.

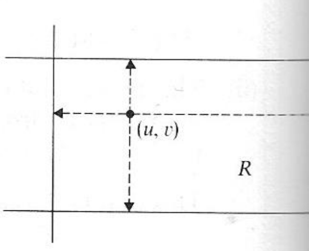


Fig. 10

Theorem 4

If A and B are open so are

(i) $A \cap B$,

(ii) $A \cup B$.

Proof

(i) If $A \cap B = \varnothing$ we already know that $A \cap B$ is open. Otherwise, let $a \in A \cap B$. Then there is δ_1 such that if $|z - a| < \delta_1$ then $z \in A$, and δ_2 such that if $|z - a| < \delta_2$ then $z \in B$. Let $\delta = \min(\delta_1, \delta_2)$. Then if $|z - a| < \delta$, we have $z \in A$ and $z \in B$, that is $z \in A \cap B$. Hence a is an interior point of $A \cap B$.

(ii) If a is an interior point of a set A it is clearly an interior point of any set containing A. Thus if A and B are open so is $A \cup B$: for if $a \in A \cup B$ it is either in A or in B—if in A, it is an interior point of A and so of $A \cup B$, and likewise if in B. ■

Each of these results can be extended by induction to any finite number $A_1, \ldots, A_n$ of open sets: thus $A_1 \cap \ldots \cap A_n$ and $A_1 \cup \ldots \cup A_n$ are open. The case of infinite intersections and unions is discussed in Problem 4 of the next section.

The majority of elementary geometric "open" figures are constructed from the sets of Example 2 by a finite number of intersections and unions: thus they are all open in the precise sense. Here are some examples.

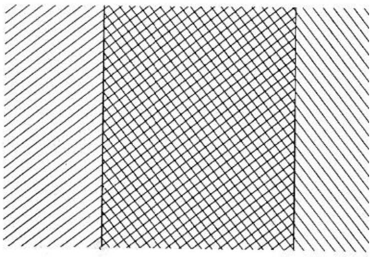


Fig. 11

Example 4

We show that the following sets are open, by building them up from simpler ones.

(i) The strip $S_1 = \{(x, y): a < x < b, y \in \mathbf{R}\}$. We can write

$$S_1 = \{(x, y): x > a\} \cap \{(x, y): x < b\}.$$

Hence S_1 is the intersection of two open half-planes (Fig. 11), and so is open.

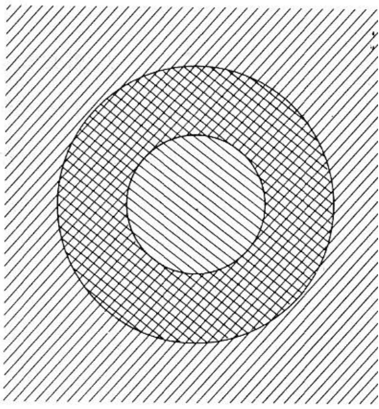


Fig. 12

(ii) The annulus $S_2 = \{z: r_1 < |z - a| < r_2\}$. We can write

$$S_2 = \{z: |z - a| > r_1\} \cap \{z: |z - a| < r_2\}.$$

Hence S_2 is the intersection of the outside of a circle and an open disc (Fig. 12), and so is open.

(iii) The sector $S_3 = \{z: 0 < |z| < 1, 0 < \text{Arg } z < \pi/2\}$. We can write

$$S_3 = \{z: |z| < 1\} \cap \{z: \text{Re } z > 0\} \cap \{z: \text{Im } z > 0\}.$$

Hence S_3 is the intersection of an open disc with two open half-planes (Fig. 13), and so is open.

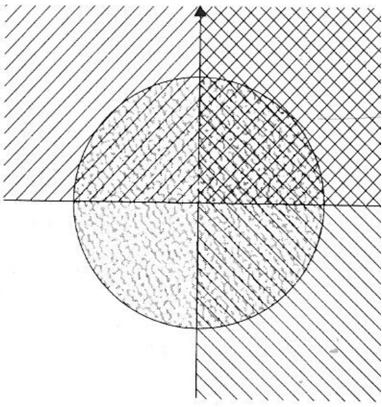


Fig. 13

There are other sets, such as parabolic or elliptical sectors, which we cannot yet prove to be open in a systematic fashion (the words "prove" and "systematic" are both important here). However, in the next reading section, when we reformulate continuity in terms of open sets, we shall be able to deal with these.

Summary

In this section we have introduced the important concepts of interior point, limit at a punctured interior point, continuity at an interior point, and open set. We have shown how to prove results on limits and continuity by using the definitions, and how limits and continuity behave under the various operations on functions.

Then we discussed open sets, and we showed how to prove that a set was open using the definition. We concluded with the result that the intersection and union of a finite number of open sets is open, and an example of how to use it.

Self-Assessment Questions

1. Which one of the following is equivalent to the definition of "f has limit l near a"?

   (i) For every $\delta > 0$ there is some $\varepsilon > 0$ such that $|f(z) - l| < \delta$ whenever $|z - a| < \varepsilon$ and $z \neq a$.

   (ii) For each $\varepsilon > 0$ there is some $\delta > 0$ such that $|f(z) - l| < \varepsilon$ whenever $0 \leqslant |z - a| < \delta$.

   (iii) For all $\varepsilon > 0$ there is some $\delta > 0$ such that if $|f(z) - l| < \varepsilon$ then $0 < |z - a| < \delta$.

2. Given that f and g are continuous at a, which of the following must be true? Give your reasons.

   (i) $z \longrightarrow [f(z)]^2 + 2g(z) + 3$ is continuous at a;

   (ii) $f/|g|$ is continuous at a;

   (iii) $g \circ f$ is continuous at a;

   (iv) $\mathrm{Re}(f + \mathrm{Im}\, g)$ is continuous at a.

3. Prove from the definition of limit that the function $f(z) = 3|z|$ is continuous at each point a of $\mathbf{C}$.

4. Let $\lim_{z \to a} f(z) = l$. Define g by

$$g(z) = \begin{cases} f(z), & \text{if } z \neq a \\ l, & \text{if } z = a. \end{cases}$$

   Show that g is continuous at a.

Solutions

1. Statement (i) is equivalent. (We interchanged ε and δ, and made a few other harmless alterations.) The others are not.

2. (i) must be true (see Theorem 2).

   (ii) can be meaningless (suppose $g(a) = 0$).

   (iii) can be false (we must have g continuous at $f(a)$—see Theorem 3).

   (iv) must be true (since Im is continuous at $g(a)$ and Re is continuous at $f(a) + \mathrm{Im}\, g(a)$, using Theorems 2 and 3).

3. $$\begin{aligned} \big|3|z| - 3|a|\big| &= 3\big||z| - |a|\big| \leqslant 3|z - a| \\ &< \varepsilon \text{ if } |z - a| < \tfrac{1}{3}\varepsilon. \end{aligned}$$

   Thus $\lim_{z \to a} 3|z| = 3|a|$, and so f is continuous at a.

4. Let $\varepsilon > 0$. There is $\delta > 0$ such that if $|z - a| < \delta$ and $z \neq a$, then $|f(z) - l| < \varepsilon$, and so $|g(z) - g(a)| < \varepsilon$.

   Also if $z = a$ then $|g(z) - g(a)| < \varepsilon$. Thus g is continuous at a.

2.4 PROBLEMS

1. Show that f is continuous at a if for every neighbourhood D of $f(a)$ there is a neighbourhood D' of a such that $f(D') \subseteq D$.

2. Suppose that $\lim_{z \to a} f(z) = l$ and g is continuous at l.

   Show that $\lim_{z \to a} g(f(z)) = g(l)$. (Thus "continuous functions preserve limits".)

3. Find the following limits:

   (i) $\lim_{z \to 2} \bar{z}^2$;

   (ii) $\lim_{z \to 1} \dfrac{z^3 + 1}{z + 1}$;

   (iii) $\lim_{z \to i} \dfrac{z^3 + i}{z - i}$.

4. You have seen that the intersection and union of any finite number of open sets is open.

   (a) Show that the intersection of an *infinite* collection of open sets is not necessarily open.

   (b) However, show that the union of any collection of open sets is open.

5. Suppose that $\lim_{z \to a} f(z) = l$. Show that for each $\varepsilon > 0$ there is $\delta > 0$ such that whenever $0 < |z_1 - a| < \delta$ and $0 < |z_2 - a| < \delta$, then $|f(z_1) - f(z_2)| < \varepsilon$. (This is similar to the Cauchy condition for limits of a sequence: see **Spivak**. It is also possible, though harder and less useful, to prove the converse.)

The result of the last problem is useful in proving that certain limits do *not* exist. The next problem illustrates this.

6. Show that the function $f(z) = \bar{z}/z$ has no limit near 0.

Solutions

1. This is a trivial rephrasing of the definition of continuity, since a neighbourhood D is specified by its radius and centre. Now f is continuous at a if and only if for every $\varepsilon > 0$ there is $\delta > 0$ such that

$$|f(z) - f(a)| < \varepsilon \text{ if } |z - a| < \delta,$$

that is, $f(z) \in \{w : |w - f(a)| < \varepsilon\}$ if $z \in \{w : |w - a| < \delta\}$.

Just take $D = \{w : |w - f(a)| < \varepsilon\}$ and $D' = \{w : |w - a| < \delta\}$.

We can represent this as follows.

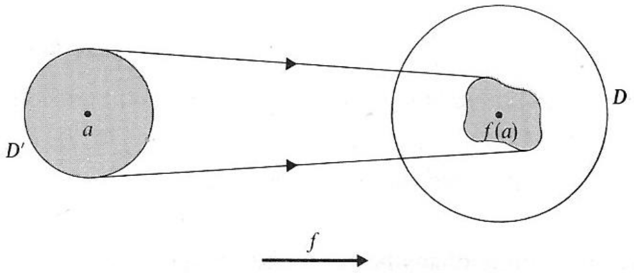


Fig. 14

2. Let $\phi(z) = \begin{cases} f(z), \text{ if } z \neq a \\ l \;, \text{ if } z = a \end{cases}$. Then ϕ is continuous at a. Since g is continuous at $l = \phi(a)$, $g \circ \phi$ is continuous at a, that is $\lim\limits_{z \to a} g(\phi(z)) = g(\phi(a)) = g(l)$, or in other words,

$$\lim_{z \to a} g(f(z)) = g(l) \quad (\text{because } \phi(z) = f(z) \text{ whenever } z \neq a).$$

3. (i) Since $z \longrightarrow \bar{z}^2$ is continuous at 2, $\lim\limits_{z \to 2} \bar{z}^2 = \bar{2}^2 = 4$.

(ii) Since $z \longrightarrow z^3 + 1$ and $z \longrightarrow z + 1$ are continuous, and $1 + 1 \neq 0$,

$$\lim_{z \to 1} \frac{z^3 + 1}{z + 1} = \frac{2}{2} = 1.$$

(iii) We can write

$$\frac{z^3 + i}{z - i} = z^2 + iz + i^2 = z^2 + iz - 1, \text{ if } z \neq i.$$

Thus

$$\lim_{z \to i} \frac{z^3 + i}{z - i} = \lim_{z \to i} (z^2 + iz - 1)$$
$$= i^2 + i \cdot i - 1 \text{ (as the function is continuous at } i),$$
$$= -3.$$

4. (a) Let $G_n = \{z : |z| < 1/n\}$, $n \in \mathbf{N}$, and let $G = \bigcap_{n \in \mathbf{N}} G_n$.

The set G_n is an open disc and so open. But

$$z \in G \Leftrightarrow |z| < \frac{1}{n} \text{ for all } n \Leftrightarrow |z| = 0;$$

that is, $G = \{0\}$. Thus G is not open.

(b) Let G be the union of a collection of open sets $\{G_j\}$. Let $a \in G$. Then a must be in at least one set in $\{G_j\}$: let it be G_k. Since G_k is open there is a neighbourhood D of a in G_k and so in G. Thus a is an interior point of G. Hence G is open. (See Problem 12 of Section 2.6 of *Unit M231 2, Functions and Graphs.*)

5. Let $\varepsilon > 0$. There is $\delta > 0$ such that whenever $0 < |z - a| < \delta$ we have $|f(z) - l| < \frac{1}{2}\varepsilon$. Hence if $0 < |z_1 - a| < \delta$ and $0 < |z_2 - a| < \delta$, then

$$|f(z_1) - f(z_2)| \leqslant |f(z_1) - l| + |l - f(z_2)| < \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon = \varepsilon.$$

6. Let $\varepsilon = \frac{1}{2}$. Then for any $\delta > 0$, the δ-neighbourhood of 0 contains points x and iy, where $x, y > 0$.

Now $f(x) = \dfrac{\bar{x}}{x} = 1$ and $f(iy) = \dfrac{\overline{iy}}{iy} = -1$; so that $|f(x) - f(iy)| = 2$. This contradicts the result of Problem 5 and so f has no limit near 0.

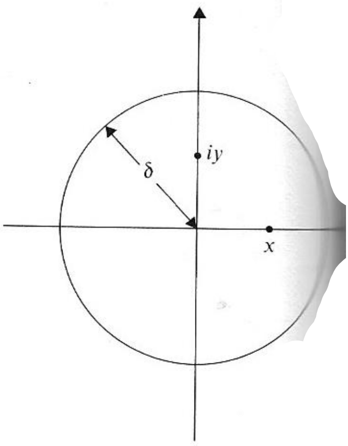


Fig. 15

2.5 CLOSED SETS

In the previous reading section, you saw how to define limits near (punctured) interior points of the domain S of some function f. Here we show how to extend the definition even further. However, we certainly cannot sensibly extend it to *all* points. Suppose, for example, that a point a is an **isolated point** of S, by which we mean that for some $\delta > 0$ the open disc of radius δ about a contains no points of S other than a (Fig. 16). Then it surely does not make sense to talk of the limit near a of f (recall that we specifically exclude $z = a$, for very good reasons, in the definition of limit). We seem to need some points of S near a; and by diminishing δ successively it becomes clear that we need some points of S as near as we like to a. This discussion should remind you of the idea of a cluster point which we met in *Unit 1*, Problem 10 of Section 1.6. We repeat that definition.

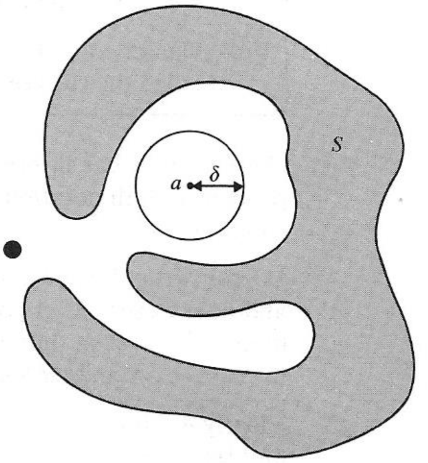


Fig. 16

Definition

Let S be any set, and a any point in **C**. Then a is called a **cluster point** of S if for all $\delta > 0$ there is a $z \in S$ such that $0 < |z - a| < \delta$. (Note that a need not belong to S.)

We shall have a look at examples of cluster points below, but before doing this, there is one observation worth making: any interior point is certainly a cluster point. However, we are more interested in cluster points not of this sort, since they provide a precise formulation of the intuitive idea of "boundary", as we shall see a little later.

Example 1

Find the cluster points of

(i) the open disc $D = \{z : |z| < 1\}$,

(ii) the closed disc $E = \{z : |z| \leqslant 1\}$.

Solution

(i) Any z with $|z| < 1$ is an interior point and so a cluster point. If $|z| > 1$ there is $\delta > 0$ such that the disc radius δ centre z does not meet D: hence z is not a cluster point. This leaves only those z with $|z| = 1$: clearly they are all cluster points.

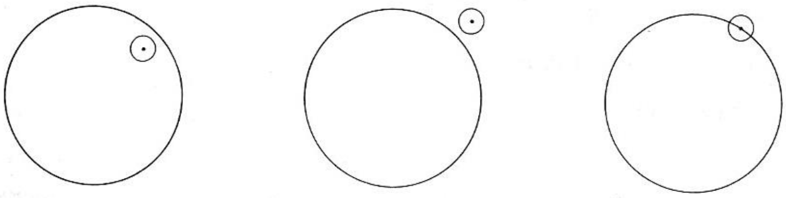


Fig. 17

(ii) Exactly the same analysis works for $E = \{z : |z| \leqslant 1\}$.

In case (ii) above E contains all of its cluster points, whereas in case (i) D does not: in fact it contains as few as possible, just the interior points. Of course, D is an open set (in the sense of Section 2.3): what do we call sets like E?

Definition

(i) A set E is called **closed** if every cluster point of E belongs to E.

(ii) The set consisting of all points of a set S and all cluster points of S is called the **closure** of S, and is denoted by $\bar{S}$.

(The use of a bar in both $\bar{z}$ (conjugate) and $\bar{S}$ (closure) should not cause any problems. Both notations are so well established that we do not wish to depart from them.)

We can write down a single condition for a point to lie in $\bar{S}$ as follows: $z \in \bar{S}$ if and only if every neighbourhood of z contains a point of S, that is, for all $\delta > 0$ there is $w \in S$ such that $|w - z| < \delta$. The easiest way to see this is to consider the cases $z \in S$ and $z \notin S$ separately.

Clearly if S is closed then $\bar{S} = S$. It is a little harder to show that the closure of any set is in fact closed—this is discussed in Problem 1 of the next section. However we shall be concerned *mainly* with closed sets rather than closure.

If you look back to Example 1, you will see that $\{z : |z| \leqslant 1\}$ is closed, while you know already that $\{z : |z| > 1\}$ is open. These sets are complements of each other. This is not an isolated example, but a general phenomenon.

Theorem 5

Let H be any set and G be its complement (that is, $G = \mathbf{C} - H$.) Then H is closed if and only if G is open.

Proof

(i) Let $a \in G$. We shall show that a is not a cluster point of H if and only if it is an interior point of G.

a is not a cluster point of H

$\Leftrightarrow$ there is $\delta > 0$ such that for all $z \in H$, $|z - a| = 0$ or $|z - a| \geqslant \delta$

$\Leftrightarrow$ there is $\delta > 0$ such that for all $z \in H$, $|z - a| \geqslant \delta$ (as $a \notin H$)

$\Leftrightarrow$ there is $\delta > 0$ such that $\{z : |z - a| < \delta\} \subseteq G$ (as $G = \mathbf{C} - H$)

$\Leftrightarrow$ a is an interior point of G.

(ii) Now G is open

$$\left.\begin{array}{l}\Leftrightarrow \text{for all } a \in G,\ a \text{ is an interior point of } G \\ \Leftrightarrow \text{for all } a \in \mathbf{C} - H,\ a \text{ is not a cluster point of } H\end{array}\right\} \text{(by (i))}$$

$\Leftrightarrow$ for all a, if $a \notin H$ then a is not a cluster point of H

$\Leftrightarrow$ for all a, if a is a cluster point of H then $a \in H$

$\Leftrightarrow$ H is closed. ■

(Sometimes, as in M231, Analysis, a set is *defined* to be closed if its complement is open, and then the *theorem* can be proved that a set is closed if and only if it contains all its cluster points. But we feel the approach here is more suitable for our purposes.)

Note that there are many sets which are neither open nor closed. A good example is the punctured disc $D = \{z : 0 < |z| \leqslant 1\}$: 0 is a cluster point of D not in D, while any point z with $|z| = 1$ is in D but is not an interior point of D (Fig. 18).

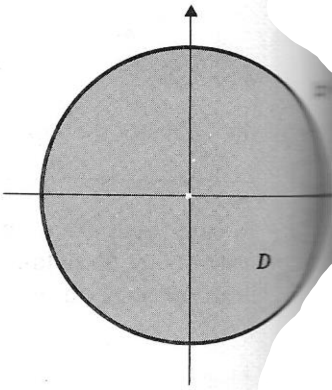


Fig. 18

In view of Theorem 5 we can translate a lot of results on open sets into results on closed sets. For example, $\varnothing$ and $\mathbf{C}$ are both closed. Most importantly, it can be shown that the intersection or union of two closed sets is closed. (See Problem 2 of the next section.)

We can then prove a lot from that result. We have already shown that the "closed disc" $\{z : |z| \leqslant 1\}$ is closed (see Example 1). Using this and other simple cases we can build up quite a stock of closed sets. In each case the "closed" sets you met in *Unit 1* turn out to be closed sets.

This sounds like a lot of work, and it would be, *if* we were actually to carry it all out. But we will not, as there is a much more efficient way of proving all the results in the preceding paragraph, and various results that we could not prove in Section 2.3, for example that a parabolic sector is open.

Theorem 6

Let f be a function continuous at each point of $\mathbf{C}$, and let S be a subset of $\mathbf{C}$. Then

(i) if S is open, then $f^{-1}(S)$ is open;

(ii) if S is closed, then $f^{-1}(S)$ is closed.

Proof

(i) Suppose that S is open. Let $a \in f^{-1}(S)$. We must show that a is an interior point of $f^{-1}(S)$. Since $f(a) \in S$, there is a neighbourhood D of $f(a)$ contained in S. But since f is continuous at a, there is a neighbourhood D' of a such that whenever $z \in D'$, $f(z) \in D$. (See Problem 1 of Section 2.4.) Thus whenever $z \in D'$, $f(z) \in S$; so, $f(D') \subseteq D \subseteq S$ and hence $D' \subseteq f^{-1}(S)$. Hence a is an interior point of $f^{-1}(S)$.

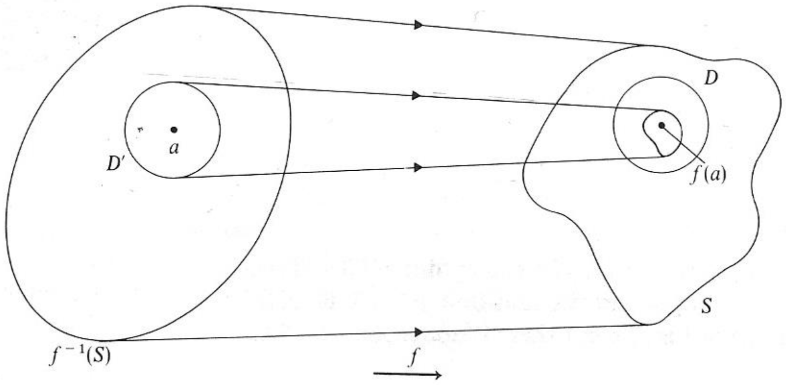


Fig. 19

(ii) Let S be closed. By Theorem 5, $\mathbf{C} - S$ is open, and so $f^{-1}(\mathbf{C} - S)$ is open, by part (i). Hence $\mathbf{C} - f^{-1}(\mathbf{C} - S)$ is closed. But

$$\begin{aligned} z \in \mathbf{C} - f^{-1}(\mathbf{C} - S) &\Leftrightarrow z \notin f^{-1}(\mathbf{C} - S) \\ &\Leftrightarrow f(z) \notin \mathbf{C} - S \\ &\Leftrightarrow f(z) \in S \\ &\Leftrightarrow z \in f^{-1}(S). \end{aligned}$$

Hence $\mathbf{C} - f^{-1}(\mathbf{C} - S) = f^{-1}(S)$, and so $f^{-1}(S)$ is closed. ∎

There is a way of modifying statements about continuous functions which often makes them of more practical use: we consider only *real-valued* functions, and interpret the appropriate notions in *real* analysis. It is also often useful to make a further reduction, replacing general sets in $\mathbf{R}$ by *intervals*.

Recall that intervals in **R** are the sets of the following forms:

$(a, b), (a, \infty), (-\infty, b), \mathbf{R}$	open intervals
$[a, b], [a, \infty), (-\infty, b], \mathbf{R}$	closed intervals
$(a, b], [a, b)$	other intervals.

(See Problem 11 of Section 2.6 of *Unit M231 2, Functions and Graphs*, for the justification of, for example, calling $[a, \infty)$ closed. See also Problem 7 of the next section.)

It is often useful to regard the empty set $\varnothing$ as an interval: it is, of course, both open and closed.

It is easy to check that an interval S is open in **R** precisely if for any $a \in S$ there is $\varepsilon > 0$ such that $(a - \varepsilon, a + \varepsilon) \subseteq S$, and closed in **R** precisely if its complement consists of one or two open intervals.

Corollary

Let f be a real-valued function continuous at each point of **C**. Then

(i) if S is an open interval of **R**, then $f^{-1}(S)$ is open;

(ii) if S is a closed interval of **R**, then $f^{-1}(S)$ is closed.

Proof

This is almost identical with that of Theorem 6. (See Problem 3 of the next section.) ■

However, the corollary is more useful than the theorem, as the next example should show.

Example 2

Show that the "open elliptical figure"

$$G = \{z : |z + a| + |z - a| < k\} \quad (\text{where } 0 < a < \tfrac{1}{2}k)$$

is open (Fig. 20), and that $F = \{z : |z + a| + |z - a| \leqslant k\}$ is closed.

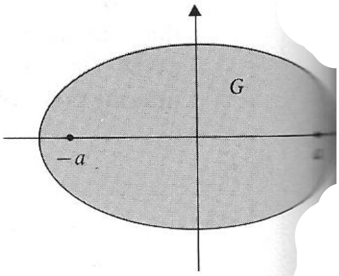


Fig. 20

Solution

Let $f(z) = |z + a| + |z - a|$. Then f is continuous (at each point of **C**) and real-valued. Let $S = (-\infty, k)$. Then S is an open interval and $G = f^{-1}(S)$. Thus G is open. (Note that although the range of f is $[2a, \infty)$, there is no reason why S has to lie in the range.)

Similarly, $F = f^{-1}([2a, k])$, and so F is closed.

There is one point we have not dealt with yet: the topological description of the "edge" or "boundary" of sets like F and G above. We define z to be a **boundary point** of a set S if z belongs to the closure of S but z is not an interior point of S. The **boundary** of a set S is the set of all boundary points of S. An open set contains none of its boundary, whereas a closed set contains all of it.

In Example 1, the boundary of the disc $\{z : |z| < 1\}$ is just the circle $\{z : |z| = 1\}$. In general, the boundary of geometrical sets is the "boundary" in the intuitive sense—for example, the boundary of G in Example 2 is just the ellipse $\{z : |z + a| + |z - a| = k\}$—but *proving* this can be a little tiresome, though not really hard. (One method is discussed in *Unit 5, Cauchy's Theorem I*.) As you may have gathered, we shall not be very concerned with proving that the intuitive boundary is the topological boundary.

We conclude by discussing limits and continuity at points which are *not* interior points. Suppose that a is a cluster point of a set S, and f is a function with domain containing S. We say that f **has limit l near a on S** if for every $\varepsilon > 0$ there is some $\delta > 0$ such that for all z, if $0 < |z - a| < \delta$ and $z \in S$, then $|f(z) - l| < \varepsilon$. (Essentially, we make the modification that the limit exists as we approach a, but only through points lying in S.)

Although this definition will be used sometimes, much more important is the definition of continuity.

Definition

Given that a is a cluster point of S, we say that f is **continuous at a on S** if f has limit $f(a)$ near a on S, that is: for every $\varepsilon > 0$ there is some $\delta > 0$ such that

$$\text{for all } z \text{ with } |z - a| < \delta \text{ and } z \in S, |f(z) - f(a)| < \varepsilon.$$

We can in fact use this definition (the ε, δ–form, not the limit form) for all points a of S. However, if $a \in S$ is not a cluster point of S, it must be an isolated point, and the definition of isolated point easily implies that f is continuous at a on S for *all* such a. (See Self-Assessment Question 6.)

We are often interested in the behaviour of the value $f(z)$ of a function f as z varies over some set S, when we do not know (or care) whether f is defined outside S. For example, consider the real-valued function $f(z) = \sqrt{1 - |z|}$ whose domain is clearly the closed disc $\{z : |z| \leqslant 1\}$. This function is continuous on S in an intuitive sense, and we would like to make this sense precise. This we do in the obvious way: we say that f is **continuous on** S if f is continuous at a on S for every $a \in S$.

In the next two reading sections we shall be very concerned with this *global* form of continuity.

Summary

In this section we have discussed cluster points and closed sets. We have proved that a set is closed if and only if its complement is open, and that the inverse image of an open (or closed) set under a continuous function is open (or closed). We concluded by discussing the definition of continuity on a set.

Self-Assessment Questions

1. Let $a \in S$. Show that a is an isolated point of S if and only if it is not a cluster point of S.
2. What are the cluster points of the set $S = \left\{\frac{1}{2^n} : n \in \mathbf{N}\right\}$?
3. Define "closed set".
4. Let f be a real-valued function, continuous at each point of $\mathbf{C}$, and let $S = (1, 2) \subset \mathbf{R}$. Why is $f^{-1}(S)$ open?
5. Show that if f is continuous at a on S and a is an interior point of S then f is continuous at a.
6. Let f be a function with domain containing S, $a \in S$ an isolated point of S. Show that f is continuous at a on S.

Solutions

1. a is not a cluster point of S

   $\Leftrightarrow$ there is some $\delta > 0$ such that for all $z \in S$, $z = a$ or $|z - a| \geqslant \delta$

   $\Leftrightarrow$ there is some $\delta > 0$ such that $\{z : |z - a| < \delta\} \cap S = \{a\}$ (as $a \in S$)

   $\Leftrightarrow$ a is an isolated point of S.

2. 0 is the only cluster point of S.

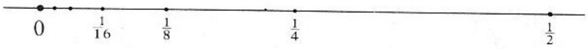


Fig. 21

3. S is closed if every cluster point of S belongs to S.

4. Because, by the Corollary to Theorem 6, the inverse image of an open interval under f is open.

5. Since a is an interior point of S, there is some $\delta_0 > 0$ such that $\{z : |z - a| < \delta_0\} \subseteq S$. Let $\varepsilon > 0$. Since f is continuous at a on S, there is some $\delta_1 > 0$ such that

   $$\text{if } |z - a| < \delta_1 \textit{ and } z \in S \text{ then } |f(z) - f(a)| < \varepsilon.$$

   Let $\delta = \min(\delta_0, \delta_1)$. Thus $z \in S$ whenever $|z - a| < \delta$. Hence if $|z - a| < \delta$ then $|f(z) - f(a)| < \varepsilon$. Thus f is continuous at a.

6. There is δ such that if $|z - a| < \delta$ and $z \in S$ then $z = a$. Hence for any $\varepsilon > 0$, if $|z - a| < \delta$ and $z \in S$ then $z = a$, and so

   $$|f(z) - f(a)| = 0 < \varepsilon.$$

   Thus f is continuous at a on S.

2.6 PROBLEMS

1. Show that the closure of a set is closed.

2. Show that if A and B are closed so are $A \cap B$ and $A \cup B$.

3. (a) Prove the Corollary to Theorem 6 (page 72).

   (b) There is another improvement of Theorem 6 that will be quite useful later in this unit. Show that if $f: G \longrightarrow \mathbf{C}$ is continuous on G and G is open, then for any open set S, $f^{-1}(S)$ is also open.

   (Hint for both parts: Adapt the proof of Theorem 6.)

4. Describe the following sets as open, closed, neither, or both. Also give a sketch of each.

   (i) $S_1 = \{z : \operatorname{Im} z > (\operatorname{Re} z)^2\}$.

   (ii) $S_2 = \{z : 1 \leqslant |z - 2| < 3\}$.

   (iii) $S_3 = \{z : (|z| + 2)^2 \geqslant 4\}$.

   (iv) $S_4 = \{z : 2 \operatorname{Re} z + 3 \operatorname{Im} z \leqslant 2\}$.

   (v) $S_5 = \{z : |z - 1| > |z + 1|\}$.

One would expect topological definitions, such as that of open set, to be invariant under isometries (for example, translations, rotations and reflections). This is so, and it can be proved in the following manner.

5. Let f be an isometry.

   (a) Show that f is one-one and that its inverse g is an isometry. (Hint: Use the characterisation of isometries given after Problem 3 of Section 2.2.)

   (b) Let G be open. Show that $f(G)$ is open.

   (c) Deduce that, if G is open, $\{z + b : z \in G\}$, $\{\bar{z} : z \in G\}$, and $\{az : z \in G\}$ (where $|a| = 1$) are open.

6. Find sets S and T and a point a such that a is a cluster point of S and a is a cluster point of T, but a is not a cluster point of $S \cap T$. (Do not spend more than a minute or two on this problem.)

7. Show that any closed interval of $\mathbf{R}$ is closed (as a subset of $\mathbf{C}$). Why is an open interval of $\mathbf{R}$ *not* open as a subset of $\mathbf{C}$?

Solutions

1. Let S be a set, $\bar{S}$ its closure. It is enough to show that any cluster point a of $\bar{S}$ is a cluster point of S, and so in $\bar{S}$. Let $\delta > 0$. There is $b \in \bar{S}$ such that $0 < |b - a| < \delta$. If $b \in S$ we are finished. If $b \notin S$ then b is a cluster point of S, and so there is $z \in S$ such that $0 < |z - b| < \delta - |b - a|$; hence $0 < |z - a| < \delta$. In either case each δ-neighbourhood of a contains a point of S, that is, a is a cluster point of S (Fig. 22).

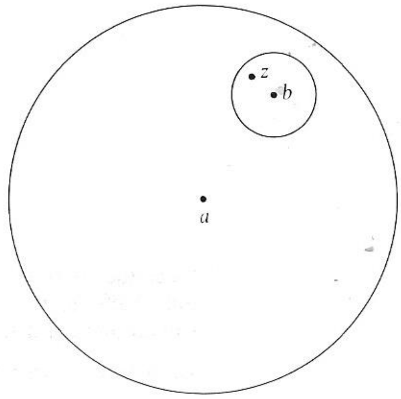


Fig. 22

2. Let A, B be closed. Then $\mathbf{C} - A$, $\mathbf{C} - B$ are open (by Theorem 5). Hence $(\mathbf{C} - A) \cap (\mathbf{C} - B)$, that is $\mathbf{C} - (A \cup B)$, is open, and so $A \cup B$ is closed. Similarly $A \cap B$ is closed.

3. (a) Given f a real-valued function continuous at each point of $\mathbf{C}$.

   (i) Let S be an open interval of $\mathbf{R}$. Let $a \in f^{-1}(S)$. Since $f(a) \in S$ there is $\varepsilon > 0$ such that

$$(f(a) - \varepsilon, f(a) + \varepsilon) \subseteq S.$$

   Since f is continuous at a, there is $\delta > 0$ such that if $|z - a| < \delta$ then $|f(z) - f(a)| < \varepsilon$, and so $f(z) \in (f(a) - \varepsilon,\ f(a) + \varepsilon)$ as $f(z) \in \mathbf{R}$, so that $f(z) \in S$. Hence $\{z : |z - a| < \delta\} \subseteq f^{-1}(S)$. Thus a is an interior point of $f^{-1}(S)$, and so $f^{-1}(S)$ is open.

(ii) Now suppose that S is a closed interval, say $S = [a, b]$. Then

$$\begin{aligned}\mathbf{C} - f^{-1}(S) = f^{-1}(\mathbf{R} - S) &= f^{-1}((-\infty, a) \cup (b, \infty)) \\ &= f^{-1}((-\infty, a)) \cup f^{-1}((b, \infty)),\end{aligned}$$

and so, by part (i), $\mathbf{C} - f^{-1}(S)$ is open, that is $f^{-1}(S)$ is closed. (The proof of the case for S a semi-infinite closed interval is similar.)

(b) Let $a \in f^{-1}(S)$. Then $f(a) \in S$ and so there is a neighbourhood D of $f(a)$ with $D \subseteq S$. Since $a \in G$ is an interior point of G, and f is continuous at a on G, f is continuous at a, and so there is a neighbourhood D' of a with $f(D') \subseteq D$. Thus $D' \subseteq f^{-1}(S)$, and so a is an interior point of $f^{-1}(S)$.

4. (i) S_1 is the inverse image of the open interval $(0, \infty)$ under the continuous function $z \longrightarrow \operatorname{Im} z - (\operatorname{Re} z)^2$ and so S_1 is open (Fig. 23).

(ii) S_2 is neither open nor closed, since $1 \in S_2$ is not an interior point of S_2 and 5 is a cluster point of S_2 not in S_2 (Fig. 24).

(iii) Since $(|z| + 2)^2 \geqslant 4 \Leftrightarrow |z| \geqslant 0$, S_3 is the whole plane $\mathbf{C}$. Hence S_3 is both open and closed.

(iv) S_4 is the inverse image of $(-\infty, 2]$ under $z \longrightarrow 2 \operatorname{Re} z + 3 \operatorname{Im} z$, and so S_4 is closed (Fig. 25).

(v) S_5 is the inverse image of $(0, \infty)$ under $z \longrightarrow |z - 1| - |z + 1|$, and so S_5 is open (Fig. 26).

Figs. 23, 24, 25 and 26 illustrate the use of thick and thin lines to indicate whether or not sets of boundary points of a set are members of that set.

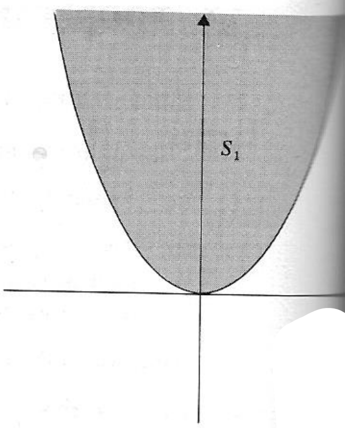


Fig. 23

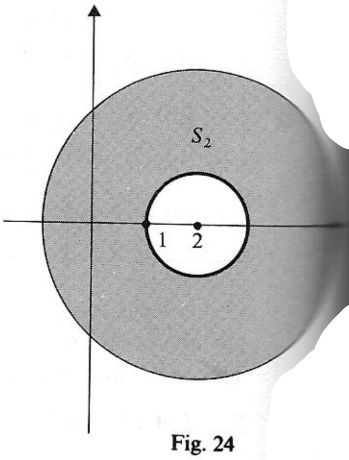


Fig. 24

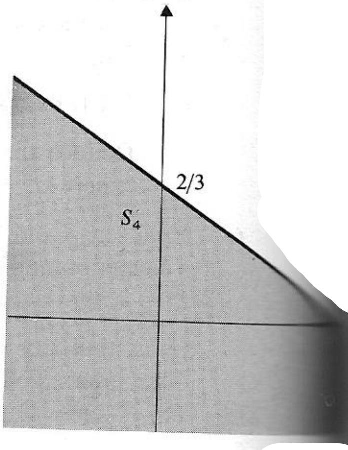


Fig. 25

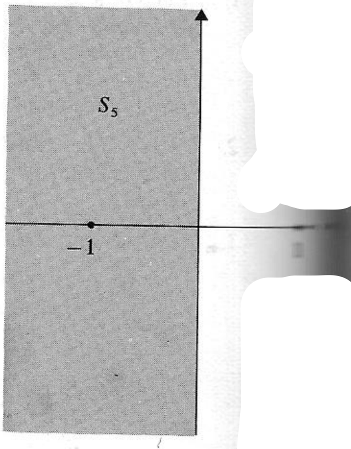


Fig. 26

5. (a) Suppose that $f(z) = az + b$, where $|a| = 1$. Now $w = az + b \Leftrightarrow z = \dfrac{1}{a}w - \dfrac{b}{a}$, and so f is one-one and the inverse g of f is also an isometry, since $\left|\dfrac{1}{a}\right| = 1$.

Similarly, if $f(z) = a\bar{z} + b$, $|a| = 1$, the inverse of f is an isometry.

(b)
$$\begin{aligned} w \in f(G) &\Leftrightarrow w = f(z) \text{ for some } z \in G \\ &\Leftrightarrow z = g(w) \text{ for some } z \in G \\ &\Leftrightarrow w \in g^{-1}(G).\end{aligned}$$

Thus, by Theorem 6, $f(G)$ is open (as g is continuous).

(c) These sets are all images of G under isometries and so, by part (b), open. In other words, open sets are preserved under translation, reflection (in the x-axis) and rotation.

(It can be deduced from this that all other topological definitions are similarly invariant.)

6. Let $S = \{(x, 0): x < 0\}$, $T = \{(x, 0): x > 0\}$. Then $(0, 0)$ is a cluster point of S and of T, but not of $S \cap T$, as $S \cap T = \varnothing$.

7. There are various ways of doing this. You could proceed directly, but we shall use Theorem 5.

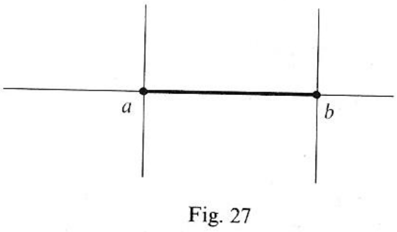


Fig. 27

Since the complement of $[a, b]$ is the union of four open half-planes, it is open. The four half-planes are $\{z: \operatorname{Re} z < a\}$, $\{z: \operatorname{Re} z > b\}$, $\{z: \operatorname{Im} z > 0\}$, $\{z: \operatorname{Im} z < 0\}$. (A similar argument shows that $[a, \infty)$, $(-\infty, b]$ and $\mathbf{R}$ are closed.)

Let S be an open interval of $\mathbf{R}$. No point of S is an interior point of S.

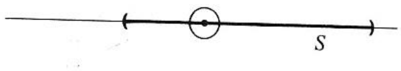


Fig. 28

2.7 CONNECTEDNESS

Now that we have studied continuous functions "locally", our next task is to find analogues of the global theorems on continuous functions of real analysis, in particular the Intermediate Value Theorem and the Boundedness Theorem (see Chapter 7 of **Spivak**). Our aim in this section is to find an analogue of the Intermediate Value Theorem.

However, this task is more general than it might appear. In some sense the two theorems above give desirable properties of intervals in **R**, and it is as a search for the correct notion of "interval" in **C** that this task should be viewed.

* * * * * * * *

The Intermediate Value Theorem of real analysis can be stated as follows: if f is continuous on an interval I, $a, b \in I$, and $f(a) < k < f(b)$, then $f(c) = k$ for some c between a and b. An equivalent version is: if f is continuous on an interval I, then $f(I)$ is an interval. (Note that a subset I of **R** is an interval if whenever a, b and c are such that $a < c < b$ and $a, b \in I$, then $c \in I$.)

Thus we want to find the most general notion of "interval" in **C** such that the image $f(I)$ under a continuous function f of an "interval" I is an "interval". However, this statement seems to define "interval" in terms of itself, and so does not look very helpful. However, we can get out of this vicious circle by the trick mentioned after Theorem 6: consider real-valued functions f such that $f(I)$ is an interval in **R**.

It is convenient (although not essential) to make the further restriction that I is an open set. Thus we can state our task as follows.

Task 1

Find a useful characterization of those open sets I in **C** such that for any real-valued function f continuous on $I, f(I)$ is an interval (of **R**).

Suppose that I is an open set and f is continuous on I, but $f(I)$ is not an interval. Then there are two points $k_1, k_2 \in f(I)$ and a point $k \notin f(I)$ such that $k_1 < k < k_2$. Let

$$G_1 = \{z \in I : f(z) < k\},\ G_2 = \{z \in I : f(z) > k\}.$$

Now G_1 and G_2 are the inverse images under f of the open intervals $(-\infty, k)$ and (k, ∞) of **R**. Thus, by the corollary to Theorem 6, G_1 and G_2 are open. Also $I = G_1 \cup G_2$, $G_1 \cap G_2 = \varnothing$ (that is G_1 and G_2 are *disjoint* sets), and yet G_1 and G_2 are both nonempty (as $k_1 \in f(G_1)$, $k_2 \in f(G_2)$). We say I is the *disjoint union* of the two nonempty open sets G_1 and G_2. (In other words, we have shown that if the range is in separate pieces so is the domain.)

Definition

Let G be an open set. We say that G is **disconnected** if there are nonempty open sets G_1 and G_2 such that

$$G_1 \cup G_2 = G, \quad G_1 \cap G_2 = \varnothing.$$

We say that G is **connected** if G is not disconnected.

The condition that the sets G_1 and G_2 are nonempty is necessary—since $\varnothing$ is open we want to exclude the trivial case $G = G \cup \varnothing$. Note also that we have defined connected (and disconnected) only for *open* sets—it could be done in general, but is not necessary for our purposes.

What we proved just before the definition can be restated as follows (the contrapositive). If an open set I is connected then the image of I under any real-valued continuous function is an interval. Conversely, it can be shown that this condition implies that I is connected. (See Problem 5 of the next section.) More generally, the image of a set under a continuous function is not more fragmented than the set itself.

It is usually fairly easy to determine if an open set G is disconnected—obviously it is if it consists of two or more "pieces" like the sets below (Fig. 29), since we just take one piece as G_1 and the rest as G_2.

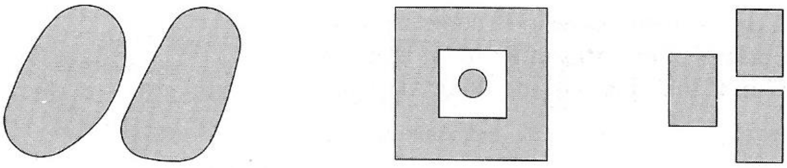


Fig. 29

However to prove that an open set G is connected you have to show that a decomposition $G = G_1 \cup G_2$ into nonempty disjoint open sets cannot occur, no matter how complicated G_1 and G_2 may be. For example, you would expect the disc $\{z : |z| < 1\}$ to be connected. Where would you put the "join" (Fig. 30)? But we cannot yet *prove* this or any other similar results.

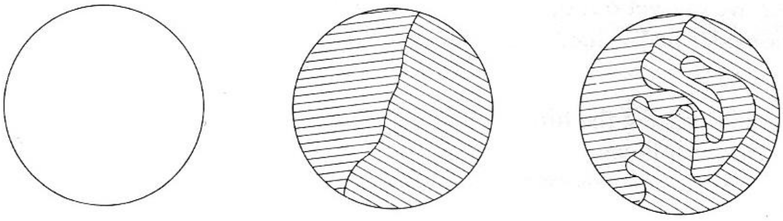


Fig. 30

As well as this, it is rather tiresome to prove, using the definition alone, that connected sets are preserved under various operations—the definition of "connected" we have, though arising naturally, is not very suitable for manipulation. In fact, although we have half-solved our task by finding a characterization, it is not very useful. Let us see if we can find an equivalent characterization which is more useful.

The sets in Fig. 31 all seem to be connected.

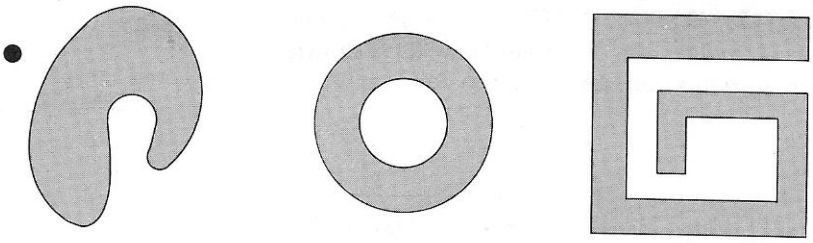


Fig. 31

If you reflect on a number of examples like these, you might come to the conclusion that in a connected set one can move from any point to any other along a continuous curve. But what *is* a continuous curve, and is this notion anyway *too* general? Well, the simplest curve is a straight line. Let us see whether we can get anywhere using straight lines to prove that certain simple open sets are connected. We shall start with a disc.

Let D be the disc $\{z : |z| < 1\}$. Suppose that D is disconnected, that is, there are nonempty open sets G_1 and G_2 disjoint from each other such that

$$G_1 \cup G_2 = D.$$

Let $z_1 \in G_1$ and $z_2 \in G_2$. We can certainly join z_1 and z_2 by a line segment lying wholly in D (Fig. 32).

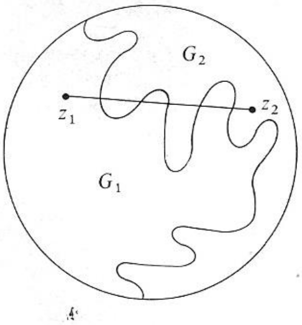


Fig. 32

Now some parts of this line segment lie in G_1 and some in G_2. The appearance of the "common boundary" of G_1 and G_2 meeting this line might look something like this:

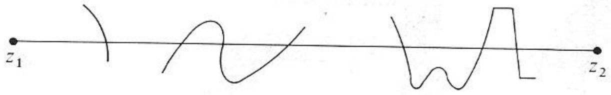


Fig. 33

If you are alert, this picture may well suggest the Intermediate Value Theorem to you (after all, connectedness is to do with this). What we shall do is to apply the technique that was used to prove this result, least upper bounds. In order to apply this, though, we shall have to interpret the problem in real analysis.

The line segment $[z_1, z_2]$ is the set of points

$$\{z : z = z_1 + t(z_2 - z_1), \quad 0 \leqslant t \leqslant 1\}.$$

In other words, it is the image of the interval $[0, 1]$ of $\mathbf{R}$ under the function γ, where

$$\gamma(t) = z_1 + t(z_2 - z_1), \quad t \in [0, 1].$$

(We shall consider such functions more fully in *Unit 4, Integration*, but we really only need $\gamma(t)$ as an abbreviation here. We refer to γ as the *standard parametrization* of $[z_1, z_2]$.) Consider then, the set

$$A = \{t \in [0, 1] : [z_1, \gamma(t)] \subseteq G_1\}$$

(just the definition you would expect from the proof of the Intermediate Value Theorem on page 113 of **Spivak**). Now A is nonempty, since $0 \in A$, and A is bounded above, since $A \subseteq [0, 1]$. Let α be the least upper bound of A. Then either $\gamma(\alpha) \in G_1$ or $\gamma(\alpha) \in G_2$.

Suppose that $\gamma(\alpha) \in G_1$. Then since G_1 is open there is $\delta > 0$ such that $\gamma(\alpha) + z \in G_1$ whenever $|z| < \delta$ (Fig. 34). In particular, this holds whenever $|z| < \delta$ *and* $z \in [z_1, z_2]$. Hence there is δ' such that whenever $0 \leqslant t < \delta'$,

$$\gamma(\alpha + t) \in G_1, \text{ that is, } \alpha + t \in A.$$

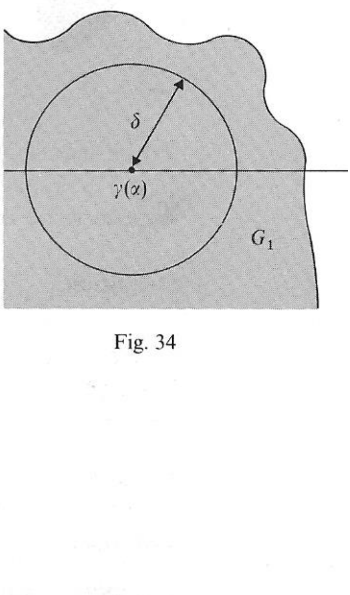


Fig. 34

Hence α is not an upper bound of A. Similarly if $\gamma(\alpha) \in G_2$ we would be able to show that α was not the *least* upper bound.

If you look closely at these arguments you will see that the first does not quite cover the case $\alpha = 1$ and the second the case $\alpha = 0$. But it is easy to show that, since z_2 and z_1 are interior points of G_2 and G_1 respectively, we have $0 < \alpha < 1$.

From the assumption that the disc D was disconnected we have arrived at a contradiction. Hence D is connected.

It is clear that this argument will work, not just for a disc, but for any open set G with the property that for any two points $z_1, z_2 \in G$, the line segment $[z_1, z_2]$ is contained in G. Such sets are called *convex*, and will be investigated much more thoroughly in *Unit 5*. Any open disc, rectangle, or half-plane (and in fact the whole plane $\mathbf{C}$) are convex. (For those people who are sticklers for rigour, we shall give analytic proofs of these geometrically obvious facts in *Unit 5*.) For the moment, though, we shall state the theorem that the last page's work has established.

Theorem 7

Any convex open set is connected.

Suppose, though, that our open set is not convex, but "bent", like the following sets.

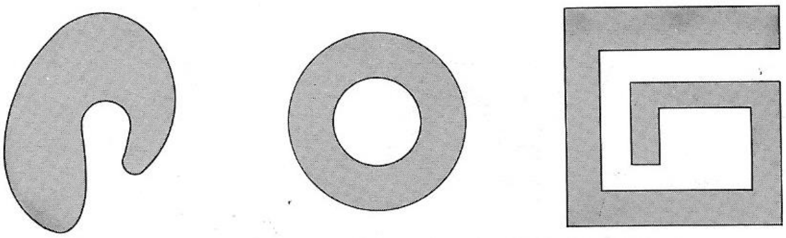


Fig. 35

We can at least join any two points z, z' in such a set by a succession $[z, z_1]$, $[z_1, z_2], \ldots, [z_{n-1}, z']$ of line segments each contained in the set (Fig. 36).

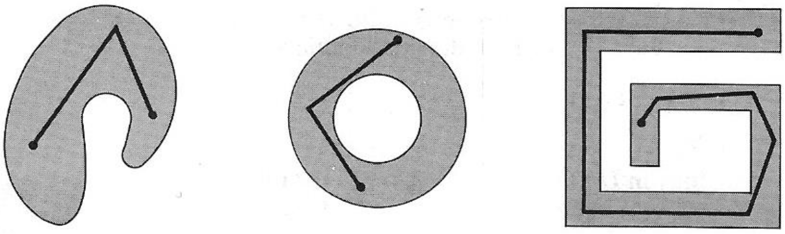


Fig. 36

To be precise, we define a **polygonal line from** z **to** z' to be a set of the form $[z, z_1] \cup [z_1, z_2] \cup \ldots \cup [z_{n-1}, z']$ where $[z, z_1], \ldots, [z_{n-1}, z']$ are line segments: we also write this set as $[z, z_1, \ldots, z_{n-1}, z']$.

Definition

We call a set G **polygonally-connected** if for any two points z, z' in G there is a polygonal line $[z, z_1, \ldots, z_{n-1}, z']$ joining them which is contained in G.

Using this notion we can get an extension of the theorem.

Corollary

Any polygonally-connected open set G is connected.

Proof

Suppose that G is disconnected. Let G_1, G_2 be two nonempty open sets such that

$$G_1 \cup G_2 = G, \quad G_1 \cap G_2 = \varnothing.$$

Let $z \in G_1$, $z' \in G_2$. Then there is a polygonal line $[z_0, z_1, \ldots, z_{n-1}, z_n]$ in G joining $z_0 = z$ and $z_n = z'$. Let k be the least number such that $z_k \in G_2$. Then $0 < k \leqslant n$ and $z_{k-1} \in G_1$. Now apply the proof of Theorem 7 to the line segment $[z_{k-1}, z_k]$ (instead of $[z_1, z_2]$ as before). ■

We can also prove the converse of the corollary and so show that any connected set is polygonally-connected. To do this we need a new idea. Given a point z_0 in an open set G it seems reasonable to say that z in G is *polygonally connected to* z_0 *in* G if there is a polygonal line in G joining z to z_0.

Let G_1 be the set of points z in G polygonally connected to z_0 in G, and G_2 be the set of points z in G not polygonally connected to z_0 in G. We show that G_1 and G_2 are open.

Lemma

G_1 and G_2 are open sets.

Proof

Let $z \in G_1$. Then, since G is open, there is $\delta > 0$ such that $w \in G$ whenever $|w - z| < \delta$. If $|w - z| < \delta$ then $[z, w] \subseteq G$, and since there is a polygonal line $[z_0, z_1, \ldots, z_{n-1}, z] \subseteq G$ the line $[z_0, z_1, \ldots, z_{n-1}, z, w] \subseteq G$: thus $w \in G_1$. Hence z is an interior point of G_1. As this holds for each $z \in G_1$, G_1 is open.

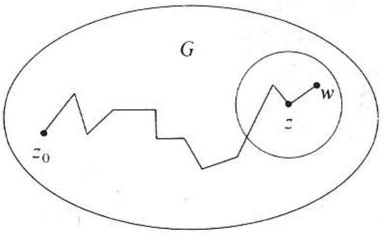


Fig. 37

Let $z \in G_2$. Again there is $\delta > 0$ such that $w \in G$ whenever $|w - z| < \delta$. Let $|w - z| < \delta$. If z_0 were polygonally connected to w in G, then z_0 would be polygonally connected to z in G (as $[w, z] \subseteq G$), hence w is not in G_1 and so must be in G_2. Thus G_2 is also open. ∎

Theorem 8

Let G be a connected open set. Then G is polygonally-connected.

Proof

Fix $z_0 \in G$. Let G_1 and G_2 be as above. Then $G = G_1 \cup G_2$, $G_1 \cap G_2 = \varnothing$, and G_1, G_2 are both open (by the lemma). As G is connected and $G_1 \neq \varnothing$, $G_2 = \varnothing$. Thus $G = G_1$ and so every point of G is polygonally connected to z_0. Since this holds for any z_0, G is polygonally-connected. ∎

It is also possible to prove from the lemma that any open set can be decomposed into a countable number of disjoint connected open sets. This result is worth remembering—but as the proof is a little off the point we have omitted it.

The notion of "connected open set" is extremely important in complex analysis, and so we give it a name.

Definition

A connected open set is called a **region**.

The work we did at the beginning of the section shows that the image of a region under a continuous real-valued function is an interval. More generally, the image of a region under *any* continuous function is a region (provided the image is open)—see Problem 1 of the next section. These two facts suggest that regions are to the complex analyst what open intervals are to the real analyst.

We finish with two examples showing how easy the formulation of connectedness in terms of polygonal connectedness is to work with. We have proved that some simple open sets (such as convex ones) are connected. More complicated sets can often be proved connected by explicit construction of polygonal lines.

Example 1

Show that the open set G, the outside of the circle $\{z:|z| = 1\}$, is connected.

Solution

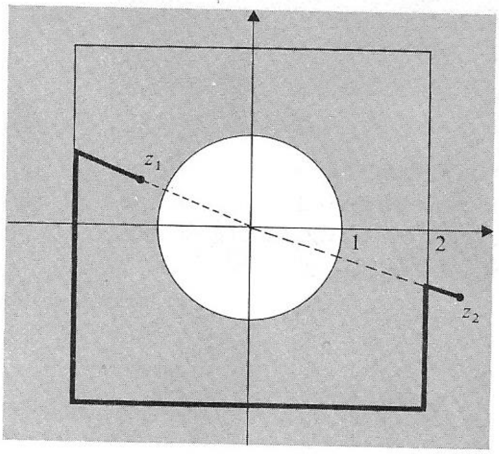


Fig. 38

Let $z_1, z_2 \in G$. Then we can join z_1 and z_2 to the outer square shown by radial lines. Hence z_1 is polygonally connected to z_2.

However, we should also investigate how connectedness behaves under operations on open sets: not very well, is one's first thought. Certainly it is possible for sets R_1 and R_2 to be connected with $R_1 \cap R_2$ not connected (Fig. 39).

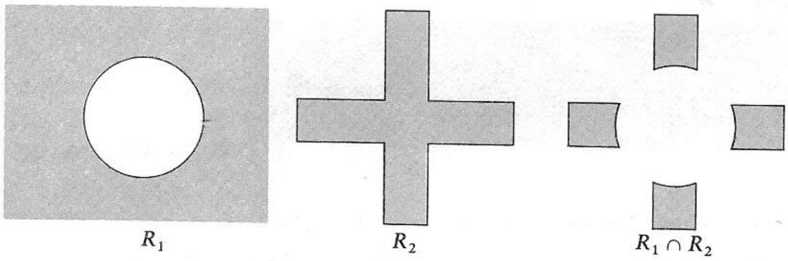


Fig. 39

Also if R_1 and R_2 are disjoint then $R_1 \cup R_2$ is certainly not connected.

Example 2

Show that if R_1, R_2 are regions and $R_1 \cap R_2 \neq \varnothing$, then $R_1 \cup R_2$ is a region.

Solution

Certainly $R_1 \cup R_2$ is open. To prove it is connected we use polygonal lines through the "overlap" $R_1 \cap R_2$. More precisely, fix $w \in R_1 \cap R_2$. If $z_1 \in R_1$ and $z_2 \in R_2$ then there is a polygonal line in R_1 from z_1 to w and one in R_2 from w to z_2: then there is a polygonal line in $R_1 \cup R_2$ from z_1 to z_2. Hence $R_1 \cup R_2$ is polygonally-connected.

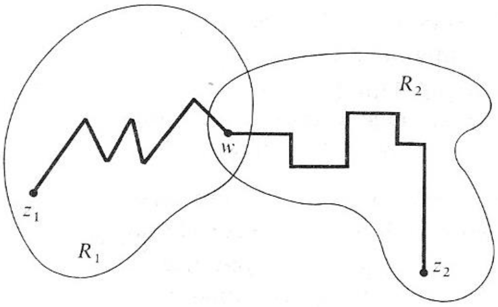


Fig. 40

The lack of a general result for $R_1 \cap R_2$ would seem to make it difficult to prove in a systematic way that standard sorts of open set are connected, and we would seem to be forced to use techniques like that in Example 1. However, in *Unit 5* we shall investigate convexity and related properties more thoroughly, and find *other* general ways of proving connectedness. For this reason we shall not spend very much time here proving that certain sets *are* regions.

Summary

In this section we have defined the notions of region (connected open set) and polygonally-connected open set, and proved them to be equivalent. We have shown that polygonal-connectedness can be verified explicitly from the definition for certain simple open sets, thus proving them connected, and that more complicated connected open sets can be built up by non-disjoint unions. The reason for introducing regions was the observation that a sort of Intermediate Value Theorem held for them, in that the image of a region under a continuous real-valued function is an interval.

Self-Assessment Questions

1. State the main steps in the proof that a convex open set is connected.
2. Let G be an open set. Define a relation $\sim$ on G by $z_1 \sim z_2$ if z_1 is polygonally connected to z_2 in G. Show, by means of figures, that $\sim$ is an equivalence relation, that is:
   (i) $z \sim z$, for all $z \in G$;
   (ii) $z_1 \sim z_2 \Leftrightarrow z_2 \sim z_1$, for all $z_1, z_2 \in G$;
   (iii) $z_1 \sim z_2$ and $z_2 \sim z_3 \Rightarrow z_1 \sim z_3$, for all $z_1, z_2, z_3 \in G$.
3. In the notation of the last question, summarize the proof that a connected open set is polygonally-connected.
4. Decide (but do not prove) which of the following sets are regions.
   (i) $\{z : |z - i| \neq |z|\}$.
   (ii) $\{z : 0 < |z| < 3\}$.
   (iii) $\{z : |\mathrm{Im}\, z| > (\mathrm{Re}\, z)^2\}$.
   (iv) $\{z : a \leqslant \mathrm{Re}\, z < b, c < \mathrm{Im}\, z < d\}$.
   (v) $\{z : 2 < |z| < 3\}$.
   (vi) $\{z : \mathrm{Re}\, z > 0 \text{ or } \mathrm{Im}\, z \neq 0\}$.
5. What can you say about the image of a region R under a real-valued function continuous on R?
6. Why is it easy to prove that the set in Fig. 41 is connected?

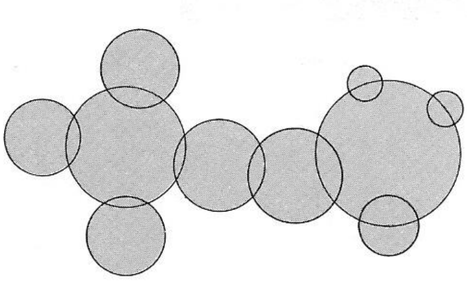


Fig. 41

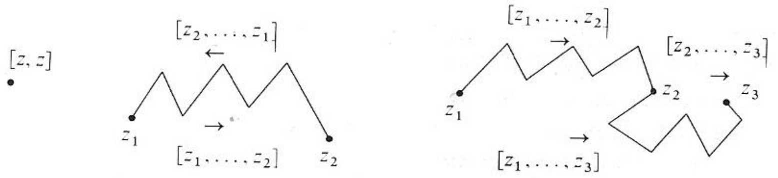
Solutions

1. (i) Using proof by contradiction, we suppose that G is a disconnected convex set. So there are nonempty open sets G_1, G_2, which are disjoint and whose union is G.

   (ii) Take $z_1 \in G_1$, $z_2 \in G_2$ and consider the line segment $[z_1, z_2]$ joining them: let γ be the standard parametrization of $[z_1, z_2]$.

   (iii) Consider the least upper bound α of those t such that $[z_1, \gamma(t)] \subseteq G_1$. Then either $\gamma(\alpha) \in G_1$ or $\gamma(\alpha) \in G_2$. Eliminate all possible values of α, to obtain a contradiction in both cases.

2. These are obvious (from the figure).



$[z, z]$
$[z_2, \ldots, z_1]$
$\leftarrow$
$\rightarrow$
z_1
$[z_1, \ldots, z_2]$
z_2
$[z_1, \ldots, z_2]$
$\rightarrow$
$[z_2, \ldots, z_3]$
$\rightarrow$
z_1
z_2
z_3
$\rightarrow$
$[z_1, \ldots, z_3]$

Fig. 42

3. Let G be a connected open set, and let $z_0 \in G$. Then $G_1 = \{z \in G : z \sim z_0\}$ and $G_2 = \{z \in G : z \nsim z_0\}$ are both open, and they are disjoint with union G. Since $z_0 \in G_1$, G_1 is nonempty. Therefore, since G is connected, $G_2 = \varnothing$ and $G_1 = G$, so G is polygonally connected.

4. (i) Open, but not a region.

   (ii) Region (see Problem 3 of the next section).

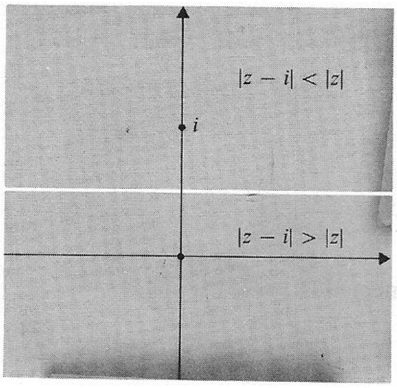


Fig. 43

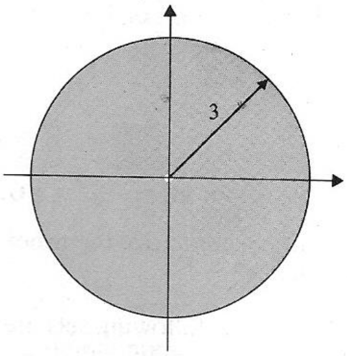


Fig. 44

   (iii) Open, but not a region.

   (iv) Not open and so not a region.

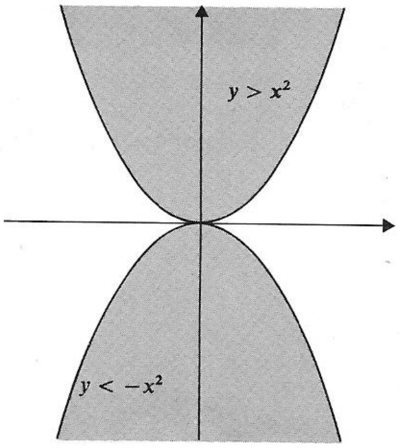


Fig. 45

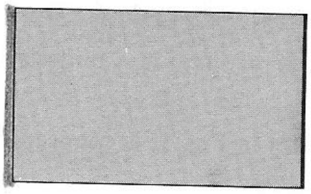


Fig. 46

(v) Region.

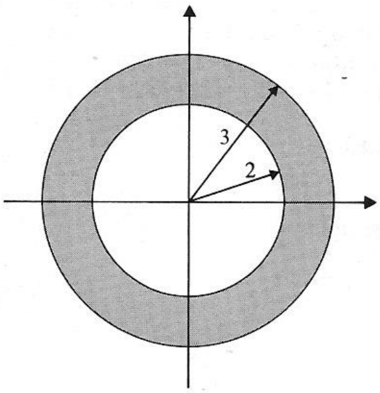


Fig. 47

(vi) Region. (See Problem 3(ii) of Section 2.8.)

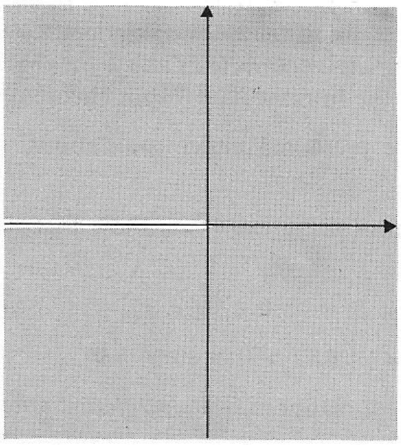


Fig. 48

5. It is an interval.

6. Each disc is connected. Now apply Example 2 repeatedly.

2.8 PROBLEMS

1. Let f be a function continuous on a connected open set R. Suppose that $f(R)$ is open. Show that $f(R)$ is connected. (This is the analogue for functions with complex range of the Intermediate Value Theorem.)

2. Show that regions are preserved under translations, rotations and reflections.

3. By actual construction of polygonal lines, show that the following open sets are connected.

   (i) The punctured disc $D = \{z : 0 < |z| < 1\}$.

   (ii) The cut plane $E = \{z : \text{Re}\, z > 0 \text{ or } \text{Im}\, z \neq 0\}$.

   (We shall be studying cut planes in *Unit 3, Differentiation*.)

4. We shall say that f is **locally constant at** a if there is a neighbourhood D of a such that

   $$f(z) = f(a) \text{ for all } z \in D.$$

   Show that if f is locally constant at each point of a region R then f is constant on R. (This result will be used several times throughout the course. It is a typical example of how connectedness is used to prove results.)

5. Suppose that G is an open set such that the image $f(G)$ of G under any real-valued function f continuous on G is an interval. Show that G is connected. (Hint: Argue by contradiction.)

Solutions

1. Suppose that E_1, E_2 are open subsets of $f(R)$ with $E_1 \cap E_2 = \varnothing$, $f(R) = E_1 \cup E_2$. Let $G_1 = f^{-1}(E_1) \cap R$, $G_2 = f^{-1}(E_2) \cap R$. Then G_1 and G_2 are open. It is easy to check that $G_1 \cap G_2 = \varnothing$ and $G_1 \cup G_2 = R$. Since R is connected, one of G_1 and G_2 is empty, so that one of E_1 and E_2 is empty. Thus $f(R)$ is connected.

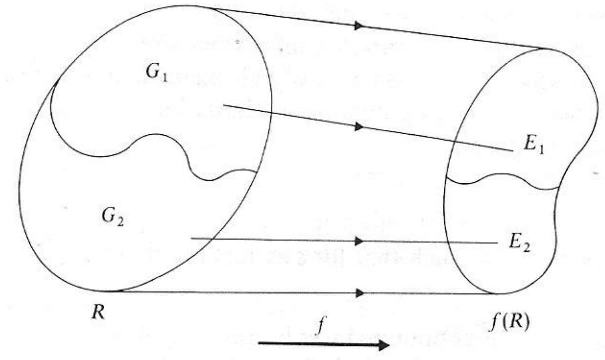


Fig. 49

2. By Problem 5 of Section 2.6, any isometry preserves open sets. By Problem 1 above any isometry preserves connected sets.

3. (i) Let $z_1, z_2 \in D$. If $[z_1, z_2] \subseteq D$, we are finished. If not, $0 \in [z_1, z_2]$. Then choose z_3 such that $[0, z_3]$ is perpendicular to $[z_1, z_2]$ [illegible] $|z_3| = \frac{1}{2}$, say (Fig. 50). Then $[z_1, z_3, z_2] \subseteq D$.

   (ii) Given $z_1, z_2 \in E$, $[z_1, 1, z_2] \subseteq E$ (Fig. 51). (We shall investigate regions which have a point like 1 in *Unit 5*.)

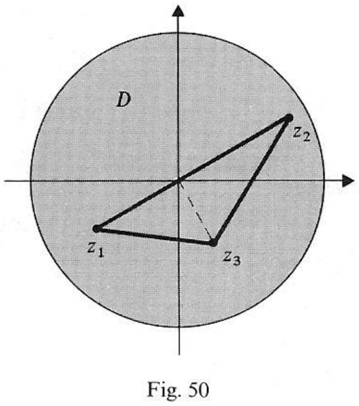


Fig. 50

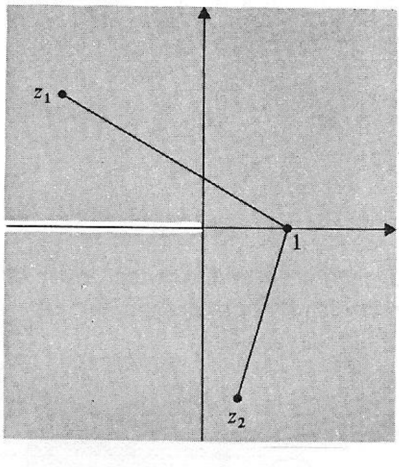


Fig. 51

4. Let $a \in R$, and consider $G_1 = \{z \in R : f(z) = f(a)\}$, $G_2 = \{z \in R : f(z) \neq f(a)\}$. Now G_2 is the inverse image of $\{w : w \neq f(a)\}$, which is an open set, and so G_2 is open (see Theorem 6). If $z \in G_1$ then some neighbourhood of z lies in G_1 (as f is locally constant at z); hence G_1 is also open. As R is connected and $G_1 \neq \varnothing$, we have $G_2 = \varnothing$, and so $G_1 = R$.

5. Let G_1, G_2 be nonempty disjoint open sets such that $G_1 \cup G_2 = G$. We define $f : G \longrightarrow \mathbf{R}$ by

$$f(z) = \begin{cases} 0, & \text{if } z \in G_1 \\ 1, & \text{if } z \in G_2. \end{cases}$$

Now we show that f is continuous on G. Let $a \in G$; we can in fact assume that $a \in G_1$, as the case for $a \in G_2$ is similar. Let $\varepsilon > 0$; we can assume that $\varepsilon \leqslant \frac{1}{2}$. Then the inverse image under f of the disc $\{z : |z - 0| < \varepsilon\}$ is just G_1, and so there is $\delta > 0$ such that $\{z : |z - a| < \delta\} \subseteq G_1$; so that $|f(z) - 0| = 0 < \varepsilon$ whenever $|z - a| < \delta$. Thus f is continuous on G.

Although f is continuous and real-valued, the image $f(G)$ of G under f is the set $\{0, 1\}$, which is not an interval. This is a contradiction.

2.9 COMPACTNESS

The other main theorem on global properties of continuous functions in real analysis is the Boundedness Theorem: if f is a function continuous on a closed interval $[a, b]$ then f is bounded on $[a, b]$.

Recall that f is bounded on $[a, b]$ precisely if there is $k > 0$ such that $|f(x)| \leqslant k$ for all $x \in [a, b]$. This definition readily generalizes to complex analysis: we say that a function f is **bounded on a set** S (contained in the domain of f) if there is $k > 0$ such that $|f(z)| \leqslant k$ for all $z \in S$. Hence in order to generalize the Boundedness Theorem we must carry out the following task.

Task 2

Find a useful characterization of those sets S in **C** such that for any function f continuous on S, f is bounded on S.

What do we know already in complex analysis about boundedness of continuous functions? If we look back to real analysis, we find one *local* result which readily generalizes. To express it we make the following definition: f is **locally bounded at** a **on** S if there is some $k > 0$ and an open disc D, centre a, such that $|f(z)| \leqslant k$ whenever $z \in D \cap S$.

Theorem 9

Let f be continuous at a on S. Then f is locally bounded at a on S.

Proof

In the definition of "f is continuous at a on S" let $\varepsilon = 1$: then there is $\delta > 0$ such that whenever $|z - a| < \delta$ and $z \in S$,

$$|f(z) - f(a)| < 1,$$

and so

$$|f(z)| < 1 + |f(a)|.$$

Then let $k = 1 + |f(a)|$, $D = \{z : |z - a| < \delta\}$. ■

Hence if f is continuous on S, each point a of S is the centre of a little disc D on which f is bounded (more precisely, f is bounded on $D \cap S$).

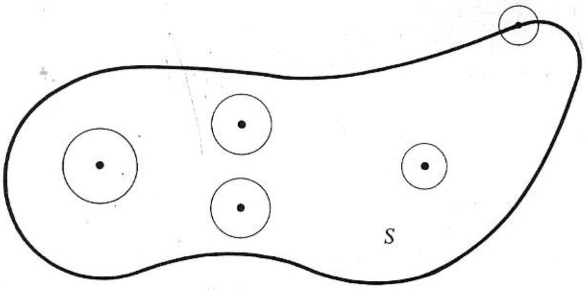


Fig. 52

If $D_1, \ldots D_n$ are such little discs, with centres $a_1, \ldots, a_n$, then, since f is bounded on each D_i, by k_i, say, it follows that f is bounded on $D_1 \cup \ldots \cup D_n$, by

$$k = \max(k_1, \ldots, k_n).$$

Note that any finite set of real numbers has a maximum (or minimum) member. But S is usually *infinite*, and so there are infinitely many discs and so infinitely many k_i – in this case we cannot be sure that there is a greatest k_i. However, each disc covers up more of S than just its centre; in fact it covers a small finite area of S. Thus it is just plausible that a finite number of them would cover all of S, if S were sufficiently "nice" and we had chosen their radii "sensibly".

If a finite number $D_1, \ldots, D_n$ cover all of S, that is, $S \subseteq D_1 \cup \ldots \cup D_n$, then f is bounded on S (by the above argument); so this is a useful property for S to have, for our purposes. Let us then formulate this "covering" property precisely, taking our courage in both hands, and boldly make a definition, not knowing yet whether it ever *can* hold.

Definition

A set S in $\mathbf{C}$ is called **compact** if no matter how an open disc D_a, centre a, is assigned to each point a of S, there is a finite number of elements $a_1, \ldots, a_n$ of S such that $S \subseteq D_{a_1} \cup \ldots \cup D_{a_n}$.

Note that each point a of S must have an open disc D_a centre a assigned to it, and that the radius of D_a can (and usually will) vary as a varies over S.

This is not the usual definition of compactness that one meets in topology (although it can be shown to be an equivalent one); however, our one turns out to be particularly suitable for complex analysis, and is (we hope) a little easier to understand than the usual topological form.

It may help to rephrase the definition as follows. A set S is compact if whenever S is covered by *any* collection $\{D_a : a \in S\}$ of open discs (with a the centre of D_a, for each $a \in S$), we can throw away all but a finite number of them and still leave S covered by the (finitely many) ones left.

Any *finite* set is compact (as then we need not throw away any discs at all). Note that finite sets can be given in somewhat disguised form—so that for example, $\{z : z^6 = 1\}$ is finite and so compact. However, we would like some infinite examples of compact sets and non-compact sets.

Example 1

Let $S = \left\{\frac{1}{2^n} : n \in \mathbf{N}\right\}$. Show that

(i) S is not compact,

(ii) $S \cup \{0\}$ is compact.

Solution

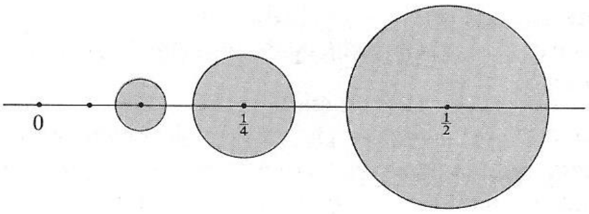
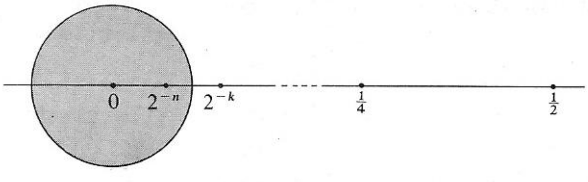


Fig. 53

(i) Let $a_n = \frac{1}{2^n}$, and let $D_{a_n} = \left\{z : \left|z - \frac{1}{2^n}\right| < \frac{1}{2^{n+2}}\right\}$. Then D_{a_n} is an open disc centre a_n, for all $n \in \mathbf{N}$. In other words, the set S is covered by the infinite collection of discs D_{a_n}. Is it possible to choose a finite number of the D_{a_n} to cover S? We shall show it is not. If $z \in D_{a_n}$ then $\frac{3}{4} \cdot \frac{1}{2^n} < |z| < \frac{5}{4} \cdot \frac{1}{2^n}$.

Hence $a_m \notin D_{a_n}$ if $m \neq n$. Thus D_{a_n} covers only a_n and no other point of S, and so, since S is infinite, no finite number of them can cover S.

(ii) Let D_{a_n} be *any* open disc centre $a_n = 1/2^n$, for $n \in \mathbf{N}$, and D_0 be *any* open disc centre 0.



The crucial difference (from (i)) is that 0 now has an open disc assigned to it. For all sufficiently large n, a_n is in D_0, since $\lim_{n\to\infty} a_n = 0$. In other words, there is some k for which $a_n \in D_0$, $n > k$. Only a finite number $a_1, \ldots, a_k$ of S are outside D_0 and we can cover these by their associated discs. Hence $S \cup \{0\}$ is covered by $k + 1$ discs: $D_0, D_{a_1}, \ldots, D_{a_k}$.

Note that the definition of "S is compact" does *not* say merely that S can be covered by *some* finite collection of open discs. Obviously S above can be covered by the *one* open disc $\{z : |z| < 2\}$, but S is not compact.

It is rather hard to prove or disprove compactness for sets using the definition alone (except in simple cases like that above, or in Problem 6 of the next section), so that we shall now derive some *properties* of compact sets.

One result is fairly obvious: a compact set is bounded. (Recall from *Unit 1* that a set S in $\mathbf{C}$ is called *bounded* if there is a closed disc D, centre the origin, such that $S \subseteq D$, or equivalently that there is a closed rectangle R such that $S \subseteq R$.)

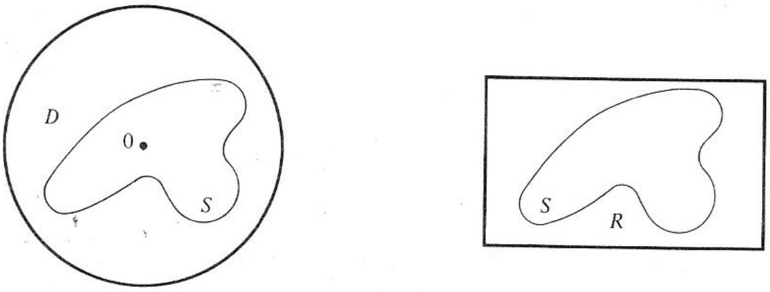


Fig. 55

For let $\{D_a : a \in S\}$ be the collection of open discs given by $D_a = \{z : |z - a| < 1\}$, $a \in S$. Then $S \subseteq D_{a_1} \cup \ldots \cup D_{a_n}$ for certain $a_1, \ldots, a_n \in S$. Let k be the maximum of $|a_j - a_1|$, $j = 1, \ldots, n$. Then for each $z \in S$,

$$\begin{aligned} |z - a_1| &\leqslant |z - a_j| + |a_j - a_1| \\ &\leqslant 1 + |a_j - a_1|, \text{ for some } j \text{ for which } z \in D_{a_j}, \\ &\leqslant 1 + k. \end{aligned}$$

Hence $|z| \leqslant |a_1| + k + 1$ for all $z \in S$.

Note that the bounded set S of Example 1 was not compact, but that the *closure* $S \cup \{0\}$ of S was compact. This might suggest the following result.

Theorem 10

Let S be a compact set. Then S is bounded and closed.

Proof

We have shown above that S is bounded. So what we have to do now is to prove that S is closed. We do this by showing that $\mathbf{C} - S$ is open, in other words that every point b of $\mathbf{C} - S$ is an interior point of $\mathbf{C} - S$. So choose any $b \in \mathbf{C} - S$. We shall construct a neighbourhood of b not meeting S. Certainly for each $a \in S$ we can take a neighbourhood E_a of b not containing a. The intersection of these E_a does not meet S – but unfortunately it will not in general be open and so may not contain a neighbourhood of b. However, if we could replace the collection of E_a by a finite number, this intersection would be open and so would contain a neighbourhood of b. To do this we use a compactness argument. But in order to use compactness we must have a neighbourhood D_a of a for each $a \in S$—and to make really sure that our eventual neighbourhood of b is disjoint from all of S we require $D_a \cap E_a = \varnothing$.

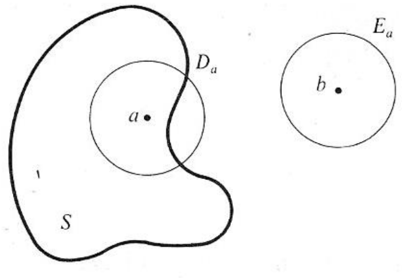


Fig. 56

Thus for each $a \in S$ let D_a be the disc radius $\frac{1}{3}|b - a|$ centre a and E_a the disc radius $\frac{1}{3}|b - a|$ centre b. Then $D_a \cap E_a = \varnothing$. (The particular radii for D_a and E_a are not important, provided that $D_a \cap E_a = \varnothing$.) Since S is compact there is a finite number $a_1, \ldots, a_n$ of points of S such that $S \subseteq D_{a_1} \cup \ldots \cup D_{a_n}$. Now $E = E_{a_1} \cap \ldots \cap E_{a_n}$ is open, so all we have to do is to show that $E \subseteq \mathbf{C} - S$. Now if $a \in S$ then $a \in D_{a_j}$ for some j and so $a \notin E_{a_j}$: thus $a \notin E$ (as $E \subseteq E_{a_j}$). Hence we have constructed an open set E containing b with $E \subseteq \mathbf{C} - S$, and so b is an interior point of $\mathbf{C} - S$. ■

You may now suspect that the converse of this theorem holds. It does, and we shall prove it. The method of proof uses a bisection argument in two dimensions, and so rests on a use of the Nested Rectangles Theorem (see Section 1.5 of *Unit 1*). You might well expect that we bisect for the property "no finite number of discs covers S", and this is exactly what we do. As long as you remember the general steps, the details should not prove too hard or confusing.

Theorem 11

Let S be a closed bounded set. Then S is compact.

Proof

Since S is bounded, we can enclose S in a closed rectangle R_1. Now let $\{D_a : a \in S\}$ be a collection of open discs with D_a centre a, $a \in S$. (For the rest of this proof we shall use "disc" to mean "disc from this collection".)

Suppose that S is not compact. Then no finite number of discs covers S. Bisect R_1 into four closed rectangles A_1, A_2, A_3, A_4 as shown (Fig. 57). If for each j some finite number k_j of discs covers $S \cap A_j$, then some finite number, at most $k_1 + k_2 + k_3 + k_4$, of discs covers S. Thus for some j (and we shall take the least one) no finite number of discs covers $S \cap A_j$. Let $R_2 = A_j$.

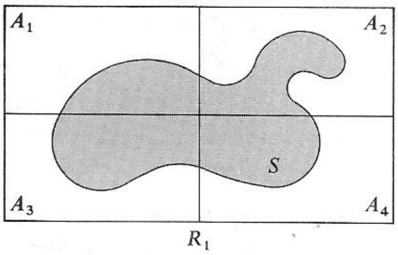


Fig. 57

We repeat this process and, by induction, construct a sequence $R_1, R_2, \ldots, R_n, \ldots$ of closed rectangles such that

(a) $R_1 \supseteq R_2 \supseteq \ldots$,

(b) $\lim_{n \to \infty} s_n = 0$ (where s_n is the sum of the lengths of the sides of R_n*: note that $s_{n+1} = \frac{1}{2}s_n$), and

(c) no finite number of discs covers $S \cap R_n$, for each $n \in \mathbf{N}$.

This is exactly the sort of situation which the Nested Rectangles Theorem was designed to deal with. Using it, there is a point ζ contained in all the rectangles R_n, $n \in \mathbf{N}$. We have now to show that $\zeta \in S$. Let $\delta > 0$: then there is an n such that $R_n \subseteq \{z : |z - \zeta| < \delta\}$; also R_n contains a point of S, in fact infinitely many (as otherwise a finite number of discs covers $S \cap R_n$). Hence ζ is a cluster point of S and since S is closed, this implies that $\zeta \in S$.

---

* In Section 1.5 of *Unit 1* we took s_n as the length of the longest side of R_n: it makes no difference.

Since $\zeta \in S$, $\zeta \in D_\zeta$. But there is some n such that R_n is contained in D_ζ. Thus $S \cap R_n$ is covered by some finite number of discs, in fact one, namely D_ζ. This is a contradiction. Hence S is compact. ■

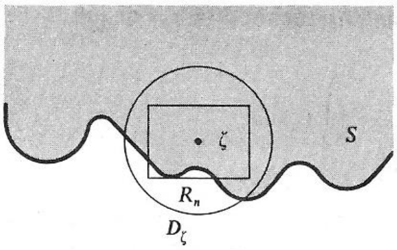


Fig. 58

This theorem is probably the hardest one in the unit, but now that we have proved it, we can deduce a large number of results by rather *easy* "compactness" arguments similar to that in Theorem 10—in fact we can sometimes get some quite surprising results by just playing around covering sets with collections of discs. Several results are given in the problems, but for now we prove the theorem which almost completes our second task.

Theorem 12 (The Boundedness Theorem)

Let S be a closed bounded set. Then any function f continuous on S is bounded on S.

Proof

Let f be continuous on S. Then by Theorem 9, for each $a \in S$ there is a neighbourhood D_a of a such that f is bounded on $D_a \cap S$. By Theorem 11, S is compact. Hence f is bounded on S, by the argument just after the proof of Theorem 9. ■

It is not hard to show that if S is a set such that any function continuous on S is bounded on S, then S is bounded and closed (see Problem 5 of the next section). Assuming this is the case our second task is complete. Note, however, that we have found a property, compactness, of very wide utility, as a by-product of our investigation.

If f is a *real*-valued function we can say a little more than in Theorem 12.

Corollary

Let S be a closed bounded set, f a real-valued function continuous on S. Then $f(S)$ has a largest and a smallest member.

Proof

By Theorem 12, f is bounded, and so $f(S)$ is a bounded set of real numbers. Let k be the least upper bound of $f(S)$. Suppose that $k \notin f(S)$. Then the function g given by

$$g(z) = \frac{1}{k - f(z)}$$

is continuous on S, and so bounded on S. However, given $\varepsilon > 0$ there is $z \in S$ such that $0 < k - f(z) < \varepsilon$, and so $g(z) > 1/\varepsilon$. Thus g is not bounded on S. This is a contradiction and so $k \in f(S)$. Let $f(z_0) = k$. Then $f(z_0) \geqslant f(z)$ for all $z \in S$—in other words, $f(z_0)$ is the largest element of $f(S)$. By applying the preceding argument to $-f$, we can prove that $f(S)$ has a smallest member. ■

There is more traditional terminology for this situation. We call the largest element of $f(S)$ the **maximum of f on S**, and denote it by $\max\{f(z): z \in S\}$: similarly the smallest member of $f(S)$ is called the **minimum of f on S**, and is denoted by $\min\{f(z): z \in S\}$. Thus the corollary tells us that a real-valued function continuous on a closed bounded set S has a maximum and minimum on S.

This result is very useful—one application is to distances from sets. Given a nonempty set A and a point z, we define the **distance from** z **to** A, written $d_A(z)$, to be the greatest lower bound of the set $\{|z - w|: w \in A\}$. This real-valued function d_A has several properties—here are some of them.

Example 2

Let A be a nonempty closed set. Show that:

(i) $d_A(z) = \min\{|z - w|: w \in A\}$;

(ii) $d_A(z) = 0$ if and only if $z \in A$;

(iii) d_A is continuous.

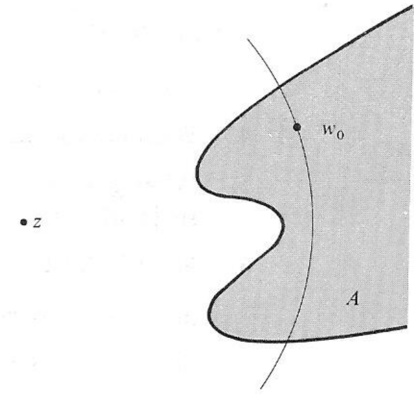


Fig. 59

Solution

(i) Let $w_0 \in A$, $r = |z - w_0|$. Then $B = A \cap \{w: |w - z| \leqslant r\}$ is a closed *bounded* set, and so $w \longrightarrow |z - w|$ has a minimum, k, on B. Thus if $w \in A$ and $|w - z| \leqslant r$ then $|w - z| \geqslant k$.

But since $k \leqslant r$, if $w \in A$ and $|w - z| > r$ then $|w - z| > k$, so that k is the minimum of $w \longrightarrow |z - w|$ on A.

(ii) If $z \in A$ then $|z - z| = 0$ and so $d_A(z) = 0$. Conversely, if $d_A(z) = 0$ then by (i), $|z - w| = 0$ for some $w \in A$; thus $z \in A$ (since $z = w$).

(iii) We shall show that $|d_A(z) - d_A(z')| \leqslant |z - z'|$. For each $w \in A$,

$$d_A(z) \leqslant |z - w| \leqslant |z - z'| + |z' - w|,$$

and so

$$d_A(z) - |z - z'| \leqslant |z' - w|.$$

Since $|z' - w| = d_A(z')$ for some $w \in A$,

$$d_A(z) - |z - z'| \leqslant d_A(z').$$

Similarly, $d_A(z') - |z' - z| \leqslant d_A(z)$, and the result follows.

Summary

In this section we have defined the notion of compact set, and proved that a set is compact if and only if it is closed and bounded. Using this notion of compactness we proved that a function continuous on a closed bounded set is bounded on the set. In several places we have used "compactness arguments" to derive results; this style of argument is important.

We finished the section by discussing maximum and minimum values of a real-valued function, and the notion of distance from a point to a set.

Self-Assessment Questions

1. Define "compact set".
2. What two properties of a set are together equivalent to compactness?
3. Which of the following sets are compact?
   (i) $\{z : \operatorname{Re} z = \operatorname{Im} z\}$.
   (ii) $\{z : z^9 + 3z^2 + 1 = 0\}$.
   (iii) $[0, 1]$.
   (iv) $\{z : |z| < 1\}$.
4. What are the main steps in the proof of Theorem 11?
5. What general result tells us that $z \longrightarrow (\operatorname{Re} z - 2)^2$ has a minimum on the set $\{z : |z| \leqslant 1\}$?
6. What is the distance from z to A when
   (i) $z = 1, \quad A = \{z : |z| < 1\}$;
   (ii) $z = 2, \quad A = \{z : |z| \leqslant 1\}$;
   (iii) $z = 0, \quad A = \{z : \operatorname{Re} z + \operatorname{Im} z = 1\}$?

Solutions

1. See page 89.
2. Closed and bounded.
3. (i) Not compact (because unbounded).
   (ii) Compact (it has at most nine points).
   (iii) Compact.
   (iv) Not compact (because not closed).
4. (i) Let S be closed and bounded, and let $\{D_a : a \in S\}$ be a collection of discs (centred at each point). Suppose that S is not compact.
   (ii) Starting with a closed rectangle R_1 containing S, construct by induction a sequence $R_1 \supseteq R_2 \supseteq \ldots$ of closed rectangles such that $\lim\limits_{n\to\infty} s_n = 0$ (where s_n is the sum of the lengths of the sides of R_n), and no finite number of discs covers $S \cap R_n$, for each n.
   (iii) Let ζ be the unique point to which $\{R_n\}$ converges. Show that $\zeta \in S$.
   (iv) Deduce a contradiction from the fact that $\zeta \in D_\zeta$.
5. Any real-valued function continuous on a closed bounded set has a maximum and minimum on that set (Corollary to Theorem 12).
6. (i) $d_A(1) = 0$ (even though $1 \notin A$; but A is open).
   (ii) $d_A(2) = 1$.
   (iii) $d_A(0) = 1/\sqrt{2}$ (Fig. 60).

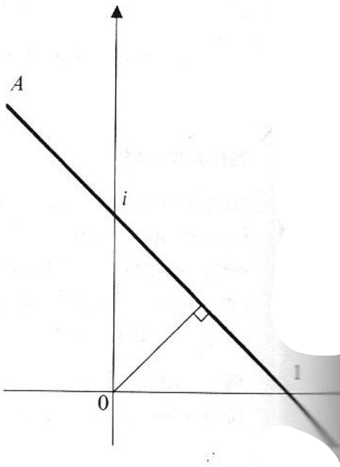


Fig. 60

2.10 PROBLEMS

The first four problems are designed to train you to use compactness arguments. Problems 1, 2 and 3 are rather general, but Problem 4 applies compactness to a specific situation.

1. Using only the definition of compactness, show that the union of two compact sets is compact.
2. Use a compactness argument to show that any closed set contained in a compact set is compact.
3. Let f be a function continuous on a compact set A. Show that $f(A)$ is compact. (In conjunction with Theorems 10 and 11 this result gives an alternative proof of Theorem 12.)

This is the general nature of compactness arguments: take the discs which some property gives you, reduce to a finite number by compactness, and see what happens. Let us try this with the example of continuity itself.

4. Suppose that f is continuous on a compact set S. Let $\varepsilon > 0$. By applying compactness to certain discs you obtain from the definition of continuity, show that there is some δ such that for all $z, z' \in S$, if $|z - z'| < \delta$ then $|f(z) - f(z')| < \varepsilon$. (Hint: Your first attempt will probably not work. Try taking the discs of half the radius that continuity implies. And have you got ε right?)

We say that a function f is **uniformly continuous on a set** S if for all $\varepsilon > 0$ there is $\delta > 0$ such that for all $z, z' \in S$, if $|z - z'| < \delta$ then $|f(z) - f(z')| < \varepsilon$. If you think about it for a moment, this says that in the definition of continuity on S, the δ for each point z, instead of depending on z, can be chosen independently of z. Problem 4 above tells us that any function continuous on a compact set S is uniformly continuous on S. We shall require this result at several points in the course (see *Unit 5*, for example).

The remaining two problems tidy up our discussion of compactness.

5. Let S be a set such that any function continuous on S is bounded on S. Show that S is (a) bounded and (b) closed. (This result finally completes our second task.)
6. Let $D = \{z : |z| < 1\}$. Using only the definition of compactness, show that D is not compact. (Hint: Make sure that D_a is well inside D, no matter where a lies in D.)

Solutions

1. Let S and T be compact. Let D_a be a neighbourhood of a, for all $a \in S \cup T$. Since S is compact there are $a_1, \dots, a_m$ in S such that $S \subseteq D_{a_1} \cup \dots \cup D_{a_m}$, and since T is compact there are $b_1, \dots, b_n$ in T such that $T \subseteq D_{b_1} \cup \dots \cup D_{b_n}$. Hence

$$S \cup T \subseteq D_{a_1} \cup \dots \cup D_{a_m} \cup D_{b_1} \cup \dots \cup D_{b_n},$$

and $\{a_1, \dots, a_m, b_1, \dots, b_n\}$ is a finite subset of $S \cup T$.

2. Let A be compact, B closed, and $B \subseteq A$. Let $\{D_b : b \in B\}$ be a collection of open discs centred on each point of B. Now $\mathbf{C} - B$ is open, and so for each point $a \in A - B$ there is an open disc D_a centre a with $D_a \subseteq \mathbf{C} - B$. Since $\{D_a : a \in B \cup (A - B)\}$ is a collection of open discs centred on each point of A, and A is compact, there are $b_1, \dots, b_m \in B$ and $a_1, \dots, a_n \in A - B$ such that

$$A \subseteq D_{b_1} \cup \ .. \cup D_{b_m} \cup D_{a_1} \cup \dots \cup D_{a_n}.$$

Since $D_{a_1} \cup \dots \cup D_{a_n}$ is disjoint from B,

$$B \subseteq D_{b_1} \cup \dots \cup D_{b_m}.$$

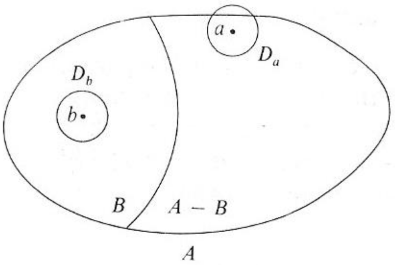


Fig. 61

3. We start off with a collection $\{D_b : b \in f(A)\}$ of open discs on $f(A)$. If we could construct from that a collection on A, we might be in business. For each $a \in A$, there is a disc D'_a centre a such that $f(D'_a \cap A) \subseteq D_{f(a)}$, since f is continuous at a on A.

Let $a_1, \dots, a_n$ be such that $A \subseteq D'_{a_1} \cup \dots \cup D_{a_n}$. Thus

$$f(A) \subseteq f(D'_{a_1} \cap A) \cup \dots \cup f(D'_{a_n} \cap A) \subseteq D_{f(a_1)} \cup \dots \cup D_{f(a_n)}.$$

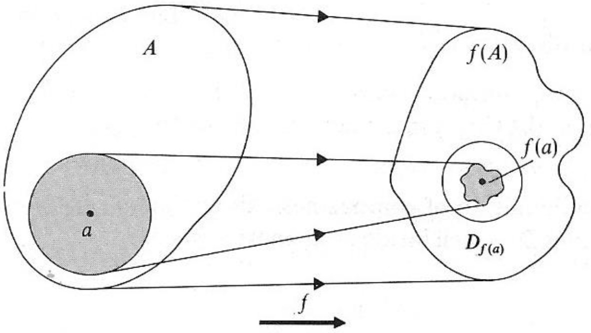


Fig. 62

4. For each $a \in S$ there is a disc D_a centre a such that if $z \in D_a \cap S$ then $|f(z) - f(a)| < \varepsilon$. By compactness of S there are $a_1, \ldots, a_n \in S$ such that

$$S \subseteq D_{a_1} \cup \ldots \cup D_{a_n}.$$

Thus if $z \in S$ then $z \in D_{a_j}$ for some j and so $|f(z) - f(a_j)| < \varepsilon$. If we could have ensured that, for z' near to z, $z' \in D_{a_j}$ also, then

$$|f(z) - f(z')| \leqslant |f(z) - f(a_j)| + |f(z') - f(a_j)| < 2\varepsilon.$$

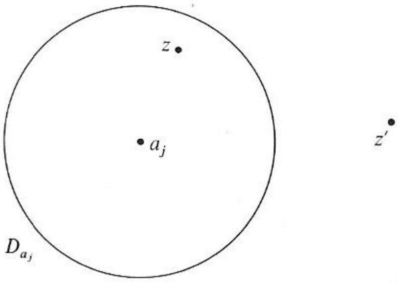


Fig. 63

So we try again. For each $a \in S$ there is $\delta_a > 0$ such that if $z \in S$ and $|z - a| < \delta_a$ then $|f(z) - f(a)| < \frac{1}{2}\varepsilon$. Let D_a be the disc centre a of radius $\frac{1}{2}\delta_a$. Then by compactness there are $a_1, \ldots, a_n \in S$ such that $S \subseteq D_{a_1} \cup \ldots \cup D_{a_n}$. Let $z \in S$. Then for some j, $z \in D_{a_j}$. But now $|z - a_j| < \frac{1}{2}\delta_{a_j}$. As long as we can ensure that $|z' - a_j| < \delta_{a_j}$ we shall be all right. This is easily done: let $\delta = \min(\frac{1}{2}\delta_{a_1}, \ldots, \frac{1}{2}\delta_{a_n})$. Now if $|z - z'| < \delta$ then $|z - z'| < \frac{1}{2}\delta_{a_j}$ and so $|z' - a_j| < \delta_{a_j}$: thus $|f(z') - f(a_j)| < \frac{1}{2}\varepsilon$ and so

$$|f(z) - f(z')| < \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon = \varepsilon.$$

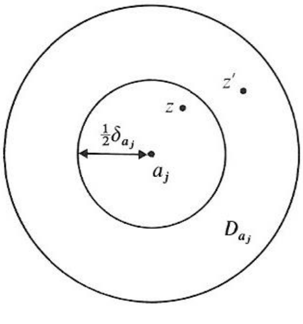


Fig. 64

5. (a) Let f be defined by $f(z) = z$, $z \in S$. Then f is continuous on S, and so bounded on S. Hence there is $m > 0$ such that $|f(z)| \leqslant m$ for any $z \in S$, that is, $|z| \leqslant m$ for any $z \in S$.

   (b) Assume that S is not closed; thus $S \neq \mathbf{C}$. Let $a \in \mathbf{C} - S$ be such that $a \in \bar{S}$. Let f be given by $f(z) = |z - a|$, $z \in S$. Then f is continuous on S. The proof of the corollary to Theorem 12 shows that f has a minimum, k, say, on S (as the proof uses only the result of Theorem 12, which is our hypothesis) and, since $f(z) \geqslant 0$ for all z, $k \geqslant 0$. But given any $\varepsilon > 0$ there is $z \in S$ such that $|z - a| < \varepsilon$ (since $a \in \bar{S}$), and so $0 \leqslant k \leqslant f(z) = |z - a| < \varepsilon$. Since $|k| < \varepsilon$ for all $\varepsilon > 0$, $k = 0$. Hence there is $z \in S$ such that $|z - a| = f(z) = k = 0$ (as f attains its minimum). Thus $z = a$ and so $a \in S$. This is a contradiction, and so S is closed.

6. We shall take D_a to be $\{z : |z - a| < \frac{1}{2}(1 - |a|)\}$, $a \in D$ (Fig. 65). Suppose that there are $a_1, \ldots, a_n \in D$ such that $D \subseteq D_{a_1} \cup \ldots \cup D_{a_n}$. Since $|a_1| < 1, \ldots, |a_n| < 1$, if we let $k = \max(|a_1|, \ldots, |a_n|)$ then $k < 1$. Thus if $1 > |a| > \frac{1}{2}(1 + k)$, then $a \notin D_{a_1} \cup \ldots \cup D_{a_n}$. This is a contradiction. Hence D is not compact.

It should be clear that this argument can be adapted to show that other similar sets—for example the punctured disc $\{z : 0 < |z| \leqslant 1\}$—are not compact.

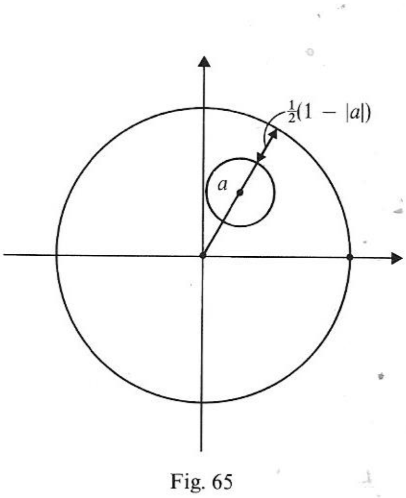


Fig. 65

Unit 3 Differentiation

Conventions

Before working through this text make sure you have read *A Guide to the Course: Complex Analysis.*

References to units of other Open University courses in mathematics take the form:

Unit M100 13, Integration II.

The set book for the course M231, Analysis, is M. Spivak, *Calculus*, paperback edition (W. A. Benjamin/Addison-Wesley, 1973). This is referred to as:

Spivak.

3.0 INTRODUCTION

In this unit we begin the study of complex analysis proper, by studying the differentiation of complex functions.

The definition of differentiability and the elementary properties of differentiable functions are very similar to the corresponding items in real analysis. The interpretation of the derivative must be rather different though: we cannot very usefully think of it as the slope of a graph, since too many dimensions are involved. Instead we revert to the other role of the real derivative, as a scale factor.

Requiring that a complex function be differentiable is a much stronger condition than you may expect from your experience of real analysis. This is manifested in two ways. Firstly, differentiable complex functions have remarkably nice properties, as will become more and more apparent as the course progresses. Secondly, the condition of differentiability is more restrictive than you might imagine. Several examples of this phenomenon occur in the present unit.

This unit is also concerned with defining new functions, or rather extending the domains of such real functions as exp, sin and cos into the complex plane.

The last two sections of the unit deal with a topic which is traditionally part of elementary complex analysis courses, the Cauchy–Riemann equations. Certain important results arise in these sections from methods involving partial derivatives of functions of two variables, but we shall see later in the course that the crucial theorems can be proved using only complex analytic techniques. This is one reason why you are required to read these two sections in a somewhat superficial way (as instructed in the sections). More significantly you may not be familiar with differentiability of functions of two variables, but it is not essential for this course that you should be. We require only that you should be aware of the significant aspects of this alternative treatment of some parts of the course. You may need to examine these sections in greater depth if you intend to study the applications of complex analysis to areas such as electrostatics, heat conduction and fluid flow.

Television

In the second television programme associated with the course we discuss the relationship between complex logarithm functions and the complex exponential function.

3.1 DERIVATIVES

This section is devoted to the definition of the derivative of a complex function, and to establishing the basic rules of differentiation. You will find that most of the results are very similar to the corresponding results in real analysis. The proofs are similar too, so we have left several of them as problems.

* * * * * * * *

We begin with some definitions.

Definition

A function f is said to be **differentiable** at an interior point a of its domain if $\lim_{z \to a} \frac{f(z) - f(a)}{z - a}$ exists. When f is differentiable at a, $\lim_{z \to a} \frac{f(z) - f(a)}{z - a}$ is called the **derivative of f at a** and denoted by $f'(a)$. The function $f': z \longrightarrow f'(z)$, whose domain is the set of points at which f is differentiable, is called the **derived function**, or the **derivative**, of f.

We have simply taken the definition of differentiability for a real function and made the appropriate changes. There are a few points worth noting. First, the limit is the limit in the sense of complex analysis: in other words, we must consider $\frac{f(z) - f(a)}{z - a}$ for *all complex numbers* z close to a. Second, if we say that a function f is differentiable at a, then there is an implicit assumption that a is an interior point of the domain of f. And third, we can define the derivative equally well by $\lim_{h \to 0} \frac{f(a + h) - f(a)}{h}$.

In complex analysis, we use a special term to describe a function differentiable on a region (that is, on an open connected set). Since this course is mainly about such functions we want you to be familiar with this term right from the start.

Definition

A function which is differentiable at every point of a region R is said to be **analytic on R**. A function analytic on **C** is said to be **entire**.

The next few results establish properties of differentiable functions and rules of differentiation that parallel those in real analysis. The proofs follow similar lines as well: see **Spivak**, Chapters 9 and 10.

Theorem 1

If f is differentiable at a, then f is continuous at a.

Proof

Since f is differentiable at a, $\lim_{h \to 0} \frac{f(a + h) - f(a)}{h}$ exists; and so does $\lim_{h \to 0} h$, of course. Thus $\lim_{h \to 0} \frac{f(a + h) - f(a)}{h} \cdot h = \lim_{h \to 0} f(a + h) - f(a)$ exists, and equals

$$\lim_{h \to 0} \frac{f(a + h) - f(a)}{h} \cdot \lim_{h \to 0} h = f'(a) \cdot 0 = 0.$$

So $\lim_{h \to 0} f(a + h) = f(a)$, and f is continuous at a. ■

Theorem 2

The functions f and g where $f(z) = c$ (a fixed complex number) and $g(z) = z$, are entire. Moreover $f'(z) = 0$ and $g'(z) = 1$ for all z.

Proof

$\dfrac{f(a+h) - f(a)}{h} = \dfrac{c - c}{h} = 0$, so f is differentiable at a for every a, and $f'(a) = 0$.

The proof for g is left to you. (See Self-Assessment Question 1.) ■

Theorem 3

If f and g are differentiable at a, then so are $f + g$, $f \cdot g$, and provided that $g(a) \neq 0$, so is f/g. Their derivatives are:

$$(f + g)'(a) = f'(a) + g'(a);$$

$$(f \cdot g)'(a) = f'(a) \cdot g(a) + f(a) \cdot g'(a);$$

$$(f/g)'(a) = [f'(a) \cdot g(a) - f(a) \cdot g'(a)]/[g(a)]^2.$$

Proof

We deal with the product, and leave the rest to you. (See Problem 1 of Section 3.3.)

$$\frac{(f \cdot g)(a+h) - (f \cdot g)(a)}{h} = \frac{f(a+h) \cdot g(a+h) - f(a) \cdot g(a)}{h}$$

$$= \frac{g(a+h)[f(a+h) - f(a)]}{h} + \frac{f(a)[g(a+h) - g(a)]}{h}.$$

Thus, using the fact that g is continuous at a,

$$\lim_{h \to 0} \frac{(f \cdot g)(a+h) - (f \cdot g)(a)}{h} = f'(a) \cdot g(a) + f(a) \cdot g'(a).$$ ■

Theorem 4

Any polynomial function is entire. Any rational function p/q, where p and q are polynomial functions, is analytic on the region $\{z : q(z) \neq 0\}$.

Proof

The product rule used inductively shows that the function $z \longrightarrow z^n$ is entire (and has derivative $z \longrightarrow nz^{n-1}$). A polynomial is a sum of products of such functions with constant functions, and is therefore differentiable everywhere. By the quotient rule, a rational function is therefore differentiable everywhere that its denominator is not zero. ■

(The statement of this theorem begs the question "Why is the set $\{z : q(z) \neq 0\}$ a region?". Well, a polynomial, other than the zero function, has a finite number of roots; that is to say, the set $\{z : q(z) = 0\}$ contains only a finite number of points. This is because each root corresponds to a linear factor of the polynomial, and the number of linear factors of a polynomial cannot exceed its degree. So the set we are interested in consists of the complex plane with a finite number of points omitted. It is not difficult to convince oneself that such a set is a region.)

Theorem 5 (The Chain Rule)

If g is differentiable at a, and f is differentiable at $g(a)$, then $f \circ g$ is differentiable at a and $(f \circ g)'(a) = f'(g(a)) \cdot g'(a)$.

Proof

Define a function ϕ as follows:

$$\phi(h) = \begin{cases} \dfrac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)}, & \text{if } g(a+h) \neq g(a) \\ f'(g(a)) & , \text{ if } g(a+h) = g(a). \end{cases}$$

Then $\phi(0) = f'(g(a))$. Moreover, ϕ is continuous at 0. We present a somewhat sketchy proof of this fact, which you can make more precise if you want to. (The procedure is the same as in the corresponding theorem in real analysis: see **Spivak**, Theorem 10-9.) If $|h|$ is sufficiently small then $|g(a+h) - g(a)|$ is small, since g is continuous at a. Thus if $|h|$ is sufficiently small and $g(a+h) \neq g(a)$, then $\phi(h) = \dfrac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)}$ is close to $f'(g(a))$. If $g(a+h) = g(a)$ then $\phi(h)$ is actually equal to $f'(g(a))$. So ϕ is continuous at 0.

Now

$$\frac{(f \circ g)(a+h) - (f \circ g)(a)}{h} = \phi(h) \cdot \frac{g(a+h) - g(a)}{h},$$

provided $h \neq 0$. (If $g(a+h) = g(a)$, both sides are zero.) It follows that

$$\begin{aligned} \lim_{h \to 0} \frac{(f \circ g)(a+h) - (f \circ g)(a)}{h} &= \lim_{h \to 0} \phi(h) \cdot \frac{g(a+h) - g(a)}{h} \\ &= f'(g(a)) \cdot g'(a). \quad \blacksquare \end{aligned}$$

(The essence of the proof is that

$$\frac{(f \circ g)(a+h) - (f \circ g)(a)}{h} = \frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)} \cdot \frac{g(a+h) - g(a)}{h}$$

if $g(a+h) \neq g(a)$. The function ϕ is brought into the argument to get round the fact that $g(a+h)$ may be equal to $g(a)$.)

These rules and results enable one to calculate derivatives of many functions, very much as in real analysis. As an example, we shall calculate the derivative of $z \longrightarrow [(z^2+1)^2+1]^2$; but before we do so, we shall dispose of a small problem of notation. The notation we have used so far would require us to write, for the derivative of the function $z \longrightarrow [(z^2+1)^2+1]^2$, something like

f', where $f(z) = [(z^2+1)^2+1]^2$,

or perhaps

$(z \longrightarrow [(z^2+1)^2+1]^2)'$.

These expressions are both rather cumbersome. So we shall often use Leibniz notation: we write $\dfrac{d}{dz}[(z^2+1)^2+1]^2$ for the derived function of this function, for example.

Example

What is the derived function of the function $z \longrightarrow [(z^2+1)^2+1]^2$?

Solution

$$\begin{aligned} \frac{d}{dz}[(z^2+1)^2+1]^2 &= 2[(z^2+1)^2+1] \cdot \frac{d}{dz}[(z^2+1)^2+1] \\ &= 2[(z^2+1)^2+1] \cdot 2(z^2+1) \cdot \frac{d}{dz}(z^2+1) \\ &= 2[(z^2+1)^2+1] \cdot 2(z^2+1) \cdot 2z \\ &= 8z(z^2+1)[(z^2+1)^2+1]. \end{aligned}$$

Now let us have some examples of functions which are not differentiable. The statement of the next theorem may surprise you a little. Note carefully that the functions listed there, some of the most frequently encountered functions of complex analysis, are *not* differentiable. It is a common error to suppose that they are.

Theorem 6

None of the functions

(i) $z \longrightarrow \bar{z}$

(ii) $z \longrightarrow |z|$

(iii) $z \longrightarrow \operatorname{Re} z$

(iv) $z \longrightarrow \operatorname{Im} z$

is differentiable anywhere in $\mathbf{C}$.

Proof

(i) Provided $h \neq 0$, $\dfrac{\overline{z+h} - \bar{z}}{h} = \dfrac{\bar{h}}{h}$. If h is real, so that $\bar{h} = h$, then $\dfrac{\overline{z+h} - \bar{z}}{h} = 1$; but if h is imaginary, then $\bar{h} = -h$, and $\dfrac{\overline{z+h} - \bar{z}}{h} = -1$. For any $\delta > 0$, the set $\{h : 0 < |h| < \delta\}$ contains both real and imaginary numbers. It therefore contains a number h_1 for which $\dfrac{\overline{z+h_1} - \bar{z}}{h_1} = 1$, and also a number h_2 for which $\dfrac{\overline{z+h_2} - \bar{z}}{h_2} = -1$. It follows that $\dfrac{\overline{z+h} - \bar{z}}{h}$ cannot approach a limit as h approaches 0, so the function $z \longrightarrow \bar{z}$ is not differentiable anywhere.

(ii) Suppose that the function $z \longrightarrow |z|$ were differentiable at some point $a \neq 0$. Then, by the Chain Rule, the function $z \longrightarrow |z|^2$ would be differentiable at a, and its product with the function $z \longrightarrow \dfrac{1}{z}$ would also be differentiable at a (since $a \neq 0$). But $\dfrac{1}{z} \cdot |z|^2 = \dfrac{1}{z} \cdot z \cdot \bar{z} = \bar{z}$, and we know that $z \longrightarrow \bar{z}$ is not differentiable anywhere. Thus $z \longrightarrow |z|$ cannot be differentiable anywhere, except possibly at 0.
To show that $z \longrightarrow |z|$ is not differentiable at 0 either, we consider $\dfrac{|0+h| - |0|}{h}$, which is just $\dfrac{|h|}{h}$. Now if h is a positive real number, $\dfrac{|h|}{h} = 1$, while if h is a negative real number, $\dfrac{|h|}{h} = -1$. It follows, just as in part (i), that $\dfrac{|h|}{h}$ does not have a limit near 0, and so $z \longrightarrow |z|$ is not differentiable at 0.

(iii) Provided $h \neq 0$, $\dfrac{\operatorname{Re}(z+h) - \operatorname{Re} z}{h} = \dfrac{\operatorname{Re} h}{h}$. If h is real, $\dfrac{\operatorname{Re} h}{h} = 1$; whereas if h is imaginary, $\dfrac{\operatorname{Re} h}{h} = 0$. Again, the argument we used in part (i) shows that $\dfrac{\operatorname{Re}(z+h) - \operatorname{Re} z}{h}$ cannot approach a limit as h approaches 0, and so $z \longrightarrow \operatorname{Re} z$ is not differentiable anywhere.

(iv) A very similar argument shows that $z \longrightarrow \operatorname{Im} z$ is not differentiable anywhere. We leave the details to you. (See Self-Assessment Question 5.) ■

There is another useful rule of differentiation: the rule for differentiating an inverse function.

Theorem 7

Let f be a one-one function defined on a region R, and let a be an interior point of $f(R)$. Suppose that f^{-1} is continuous at a, and that f is differentiable at $f^{-1}(a)$, with $f'(f^{-1}(a)) \neq 0$. Then f^{-1} is differentiable at a and

$$(f^{-1})'(a) = \frac{1}{f'(f^{-1}(a))}.$$

Proof

Set $f^{-1}(a) = b$. If h is any non-zero complex number such that $a + h \in f(R)$, there is a unique complex number, which we denote by $\phi(h)$, such that $f(b + \phi(h)) = a + h$. The function $h \longrightarrow \phi(h)$ has the following properties:

$$\phi(h) = f^{-1}(a + h) - f^{-1}(a);$$

$$f(b + \phi(h)) - f(b) = h.$$

These are just alternative ways of expressing the definition of ϕ; but they enable us to write

$$\frac{f^{-1}(a + h) - f^{-1}(a)}{h} = \frac{\phi(h)}{f(b + \phi(h)) - f(b)}.$$

Now since f^{-1} is continuous at a, $\lim_{h \to 0} \phi(h) = 0$. Moreover, since f^{-1} is one-one, $\phi(h) \neq 0$ if $h \neq 0$. It follows that

$$\lim_{h \to 0} \frac{\phi(h)}{f(b + \phi(h)) - f(b)} = \frac{1}{f'(b)},$$

as we now show in detail.

Since $f'(b) \neq 0$, it is certainly the case that

$$\lim_{h \to 0} \frac{h}{f(b + h) - f(b)} = \frac{1}{f'(b)}.$$

Thus given $\varepsilon > 0$, there is some $\delta_1 > 0$ such that

$$\left| \frac{h}{f(b + h) - f(b)} - \frac{1}{f'(b)} \right| < \varepsilon \quad \text{whenever} \quad 0 < |h| < \delta_1.$$

Since $\lim_{h \to 0} \phi(h) = 0$, there is some $\delta > 0$ such that

$$|\phi(h)| < \delta_1 \quad \text{whenever} \quad 0 < |h| < \delta;$$

and since $\phi(h) \neq 0$ for $h \neq 0$, in fact, we have

$$0 < |\phi(h)| < \delta_1 \quad \text{whenever} \quad 0 < |h| < \delta.$$

So provided $0 < |h| < \delta$,

$$\left| \frac{\phi(h)}{f(b + \phi(h)) - f(b)} - \frac{1}{f'(b)} \right| < \varepsilon.$$

In other words,

$$\lim_{h \to 0} \frac{\phi(h)}{f(b + \phi(h)) - f(b)} = \frac{1}{f'(b)}.$$

To return to the main argument, we have shown that

$$\lim_{h \to 0} \frac{f^{-1}(a + h) - f^{-1}(a)}{h} = \frac{1}{f'(b)} = \frac{1}{f'(f^{-1}(a))}.$$

Thus f^{-1} is differentiable at a and

$$(f^{-1})'(a) = \frac{1}{f'(f^{-1}(a))}. \quad \blacksquare$$

Notice that in the statement of Theorem 7 we assumed that f^{-1} is continuous at a, and that this assumption played an important role in the proof. Thus before we can use the formula for calculating the derivative of an inverse function in any particular case, we must first check that the inverse function in question is continuous.

In real analysis the situation is different, because we can prove that the inverse of a continuous one-one function must itself be continuous. But this proof depends on the fact that a continuous one-one function must be either increasing or decreasing, and since the notions of an increasing or a decreasing function make no sense in complex analysis we cannot adapt it. In fact, it *is* true that the inverse of a one-one analytic function is continuous. We could prove this result here, but the proof is rather difficult, so we defer it. Later in the course, when we have developed much more of the theory of analytic functions, we shall be able to give a straight-forward proof.

Summary

The rules of differentiation in real analysis apply in complex analysis, too, provided the appropriate modifications are made. There are two points to be careful about. Firstly, none of the new functions of complex analysis, namely conjugation, modulus, and the real and imaginary part functions, is differentiable. And secondly, before we can apply the formula for the derivative of an inverse function, we must (at this stage) show that the inverse function is continuous.

Self-Assessment Questions

1. Show, from the definition, that the function $g: z \longrightarrow z$ is differentiable at all points of the complex plane, and that $g'(z) = 1$. What special term do we use for a function which is differentiable at every point of **C**?

2. Find the derivatives of the following functions:

   (i) $z \longrightarrow \dfrac{z^2 + 2z + i}{2iz + 1}$;

   (ii) $z \longrightarrow [(z + i)^2 + (z - i)^2]^2$;

   (iii) $z \longrightarrow \dfrac{az + b}{cz + d}$ where a, b, c and d are (complex) constants.

3. Suppose that f and g are analytic on a region R. Is $f + g$ analytic on R? What about $f \cdot g$?

4. The function f is one-one on a region R; a is an interior point of $f(R)$; f is differentiable at $f^{-1}(a)$ and $f'(f^{-1}(a)) = 0$. By using the Chain Rule, show that f^{-1} cannot be differentiable at a.

5. Show that $z \longrightarrow \operatorname{Im} z$ is not differentiable anywhere.

Solutions

1. Since $g(z) = z$, we have $\dfrac{g(z+h) - g(z)}{h} = \dfrac{(z+h) - z}{h} = 1$. Thus $\lim\limits_{h \to 0} \dfrac{g(z+h) - g(z)}{h} = 1$, and so the function is differentiable at all points of the complex plane. We call such a function entire.

2. (i)
$$\frac{d}{dz}\left(\frac{z^2 + 2z + i}{2iz + 1}\right) = \frac{(2iz+1)\cdot\frac{d}{dz}(z^2+2z+i) - (z^2+2z+i)\cdot\frac{d}{dz}(2iz+1)}{(2iz+1)^2}$$
$$= \frac{(2iz+1)(2z+2) - (z^2+2z+i)2i}{(2iz+1)^2}$$
$$= \frac{2(iz^2+z+2)}{(2iz+1)^2}.$$

(ii)
$$\frac{d}{dz}([(z+i)^2 + (z-i)^2]^2) = 2[(z+i)^2 + (z-i)^2]\cdot\frac{d}{dz}[(z+i)^2 + (z-i)^2]$$
$$= 2[(z+i)^2 + (z-i)^2][2(z+i) + 2(z-i)]$$
$$= 8z[(z+i)^2 + (z-i)^2].$$

(iii)
$$\frac{d}{dz}\left(\frac{az+b}{cz+d}\right) = \frac{a(cz+d) - c(az+b)}{(cz+d)^2}$$
$$= \frac{ad - bc}{(cz+d)^2}.$$

3. The functions f and g are analytic on R; that is, they are differentiable at every point of R. But then $f + g$ and $f \cdot g$ are also differentiable at every point of R, and so are analytic on R.

4. By definition, $f \circ f^{-1}(z) = z$. If f^{-1} were differentiable at a, then by the Chain Rule we would have
$$f'(f^{-1}(a))\cdot(f^{-1})'(a) = 1.$$
But $f'(f^{-1}(a)) = 0$, so this equation cannot hold. Thus f^{-1} is not differentiable at a.

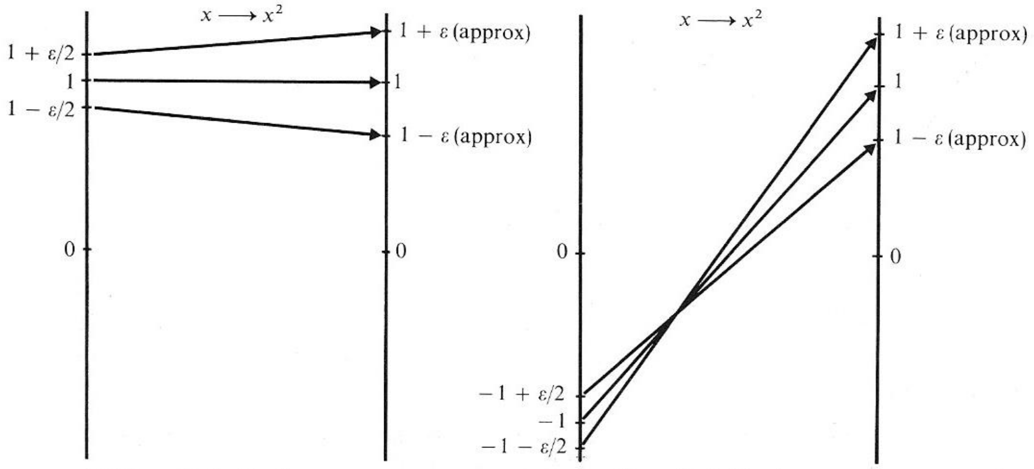
5. $\dfrac{\operatorname{Im}(z+h) - \operatorname{Im} z}{h} = \dfrac{\operatorname{Im} h}{h}$ provided $h \neq 0$. If h is real, $\dfrac{\operatorname{Im} h}{h} = 0$: if h is imaginary, $\dfrac{\operatorname{Im} h}{h} = \dfrac{1}{i} = -i$. Since every disc centred at 0 contains non-zero real and imaginary numbers, $\dfrac{\operatorname{Im}(z+h) - \operatorname{Im} z}{h}$ cannot approach a limit near 0. Thus $z \longrightarrow \operatorname{Im} z$ is not differentiable anywhere.

3.2 INTERPRETING THE DERIVATIVE

We usually think of the derivative of a real function in geometric terms—as the slope of the graph of the function. This interpretation is very useful in real analysis, but is of no use at all in complex analysis, since it is no simplification to try to picture the four-dimensional space in which the graph of a complex function must be drawn.

There is an alternative way of interpreting the derivative of a real function, which does have an analogue in complex analysis, and that is to think of it as a "scale factor".

If f is a real function which has a non-zero derivative at a point a, then the image under f of an interval $[a - \varepsilon/2, a + \varepsilon/2]$ of length ε centred at a is approximately an interval centred at $f(a)$ whose length is $|f'(a)|\varepsilon$, provided ε is sufficiently small; if $f'(a) > 0$ then the image interval has the same orientation as the original one, while if $f'(a) < 0$ the orientation is reversed. Fig. 1 illustrates the two possibilities for the function $x \longrightarrow x^2$, with $a = 1$ and $a = -1$.


$x \longrightarrow x^2$
$1 + \varepsilon/2$
1
$1 - \varepsilon/2$
0
$1 + \varepsilon$ (approx)
1
$1 - \varepsilon$ (approx)
0
$x \longrightarrow x^2$
0
$-1 + \varepsilon/2$
-1
$-1 - \varepsilon/2$
$1 + \varepsilon$ (approx)
1
$1 - \varepsilon$ (approx)
0

length multiplied by 2, orientation the same

length multiplied by 2, orientation reversed

Fig. 1

Something rather similar is true of a complex function at a point where it has a non-zero derivative; but instead of intervals we must talk of discs, and while there are only two possible orientations for an interval a disc may be rotated through any angle. The effect of a complex function f such that $f'(a) \neq 0$, when applied to a disc $|z - a| < \varepsilon$, is *approximately* to magnify or contract the disc by a factor $|f'(a)|$, and to rotate it through an angle $\operatorname{Arg} f'(a)$. Fig. 2 shows what happens for the function $f: z \longrightarrow z^2$, with $a = 1 + i$. Note that $f'(1 + i) = 2(1 + i)$, so that $|f'(1 + i)| = 2\sqrt{2}$ and $\operatorname{Arg}(f'(1 + i)) = \pi/4$, and that the approximating image disc has centre $f(1 + i) = 2i$.

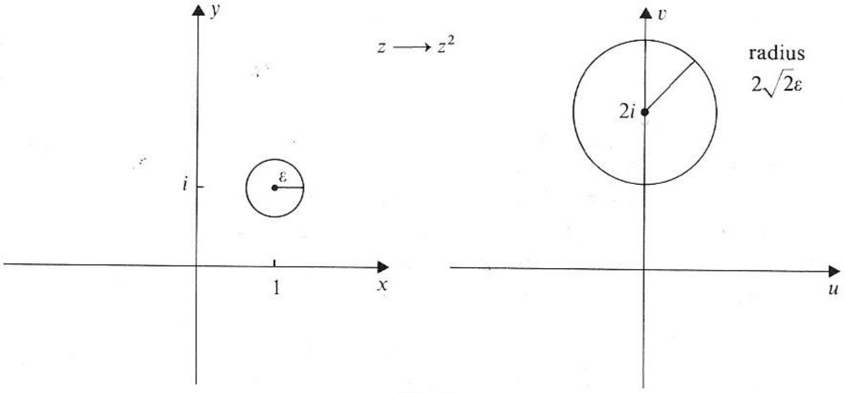

y
$z \longrightarrow z^2$
ε
i
1
x
v
radius
$2\sqrt{2}\varepsilon$
$2i$
u

Fig. 2

This interpretation of the rolè of the derivative is a consequence of the following theorem.

Theorem 8

If the function f is differentiable at a, then there is a complex function ϕ such that

$$f(z) = f(a) + f'(a)\cdot(z-a) + \phi(z) \text{ and } \lim_{z\to a}\frac{\phi(z)}{z-a} = 0.$$

Conversely, if there is a complex number ζ and complex function ϕ such that

$$f(z) = f(a) + \zeta\cdot(z-a) + \phi(z), \text{ where } \lim_{z\to a}\frac{\phi(z)}{z-a} = 0,$$

then f is differentiable at a and $f'(a) = \zeta$.

Proof

If f is differentiable at a, we can define ϕ by

$$\phi(z) = f(z) - f(a) - f'(a)\cdot(z-a).$$

Then

$$\lim_{z\to a}\frac{\phi(z)}{z-a} = \lim_{z\to a}\frac{f(z)-f(a)}{z-a} - f'(a) = 0.$$

Conversely, suppose that $\lim_{z\to a}\frac{\phi(z)}{z-a} = 0$ where

$$\phi(z) = f(z) - f(a) - \zeta\cdot(z-a).$$

Then

$$\lim_{z\to a}\frac{f(z)-f(a)}{z-a} = \lim_{z\to a}\frac{\phi(z)}{z-a} + \zeta = \zeta;$$

so f is differentiable at a, and $f'(a) = \zeta$. ■

So provided that $|z-a|$ is small, $f(z)$ is nearly the same as $f(a) + f'(a)\cdot(z-a)$. Now the effect of multiplying $z-a$ by the complex number $f'(a)$ (which we are assuming to be non-zero) is to rotate it through an angle $\operatorname{Arg} f'(a)$ and to stretch it by a factor $|f'(a)|$.

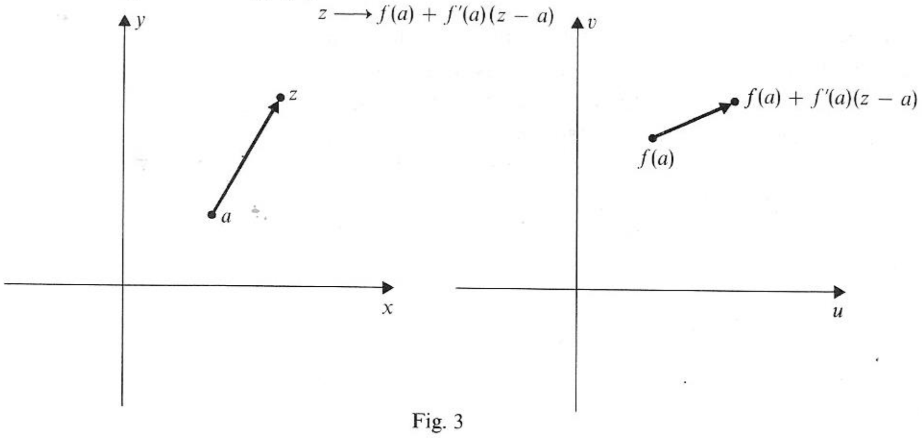


Fig. 3

The image of a disc $|z-a| < \varepsilon$ under the function $z \longrightarrow f(a) + f'(a)\cdot(z-a)$ is again a disc, centre $f(a)$, each point of which has been rotated through the angle $\operatorname{Arg} f'(a)$, and whose distance from the centre has been "stretched" by a factor $|f'(a)|$. This gives an approximation to the image of the disc under the function f, if ε is small. Thus we see how the derivative $f'(a)$ can be interpreted as a *scale factor*, causing both rotation and magnification.

To talk of approximating f by the function $z \longrightarrow f(a) + f'(a)\cdot(z-a)$, as we have been doing, is really rather unsatisfactory. It involves us in the use of vague words like "small". There is a sense in which f defines a rotation and a magnification quite precisely. The remainder of this section is devoted to describing this idea. We shall not attempt to prove any results here, because we shall be dealing with this topic in greater detail later in the course.

The idea is to look at the effect that a differentiable function has on tangents to curves through a point at which its derivative is non-zero. We have not yet defined either "curve" or "tangent" in this course; but we think that you are probably familiar enough with their meanings to follow the discussion without being given explicit definitions.

Suppose, then, that we have a curve passing through the point a, which has a tangent at a. Since we shall have to refer to the curve repeatedly, we shall denote it by a letter, the letter γ. The function f, when applied to the points of γ, generates a new curve which we denote by $f(\gamma)$. Note that $f(\gamma)$ passes through $f(a)$.

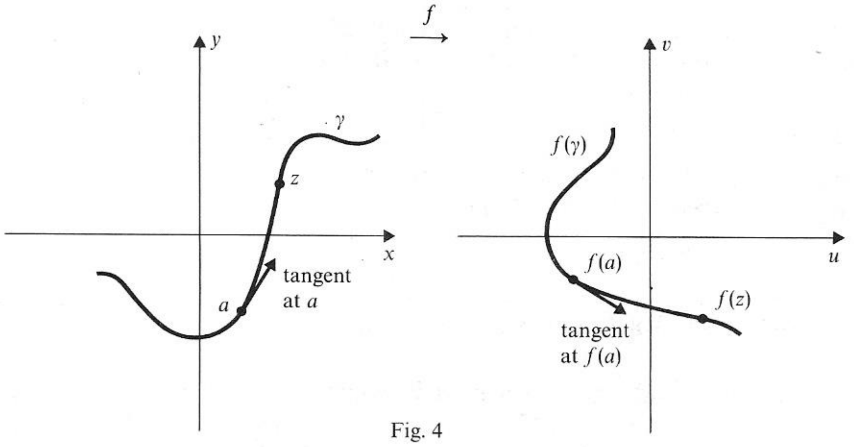


Fig. 4

The picture is supposed to clarify the definition of $f(\gamma)$: it is the set of points $\{f(z)$: the curve γ passes through $z\}$. We have assumed that γ has a tangent at a (so, for example, it does not have a "corner" there). Then $f(\gamma)$ has a tangent at $f(a)$, which may be thought of as the image of the tangent to γ at a. The tangent to $f(\gamma)$ at $f(a)$ is rotated, relative to the tangent to γ at a, through an angle $\operatorname{Arg} f'(a)$.

We sometimes think of a curve not just as a certain set of points in the plane, but as the path traced out by a moving point. The tangent is then to be interpreted as the instantaneous velocity of the moving point. As the curve γ is traced out by one moving point, a corresponding moving point traces out the curve $f(\gamma)$. We can now describe the stretching effect of the function f: in addition to the difference between the *directions* of the instantaneous velocities of the moving points on the two curves, there will be a difference in their *magnitudes*. The number $|f'(a)|$ is the ratio of the instantaneous speed (or magnitude of instantaneous velocity) of the moving point on $f(\gamma)$ as it passes through $f(a)$ to the instantaneous speed of the moving point on γ as it passes through a.

So the function f has the effect of rotating and stretching tangents to curves, or instantaneous velocities.

There is an important corollary to this result. If γ_1 and γ_2 are two curves that intersect at a, the angle between their tangents at a serves as a measure of the angle at which the curves themselves intersect. Since $f(\gamma_1)$ and $f(\gamma_2)$ intersect at $f(a)$, and the two tangents are rotated by the same amount in going from a to $f(a)$, the angle between the image curves $f(\gamma_1)$ and $f(\gamma_2)$ at $f(a)$ is the same as the angle between the curves γ_1 and γ_2 at a. In this sense,

> analytic functions preserve angles (at points at which their derivatives are non-zero).

This is an important and interesting geometric property of analytic functions.

Let us consider a special case. Suppose that γ is the line parallel to the real axis passing through the point i, in other words the set $\{z : \operatorname{Im} z = 1\}$, which may be written $\{i + t : t \in \mathbf{R}\}$. For the purposes of illustration we will choose f to be the function $f : z \longrightarrow z^3$, $z \in \mathbf{C}$. The curve $f(\gamma)$ is $\{f(i + t) : t \in \mathbf{R}\}$, which in this case becomes the set $\{(i + t)^3 : t \in \mathbf{R}\}$, and this may be written as $\{(t^3 - 3t) + (3t^2 - 1)i : t \in \mathbf{R}\}$.

The tangent to γ at i coincides with γ (since γ is a straight line), but what is the tangent to $f(\gamma)$ at $f(i) = i^3 = -i$?

We know that

$$f(i+t) = f(i) + f'(i)t + \phi(i+t)$$

(where ϕ is the remainder term defined in Theorem 8 earlier). For this particular function we therefore have

$$\begin{aligned}(i+t)^3 &= (i)^3 + 3(i)^2 t + \phi(i+t)\\ &= -i - 3t + \phi(i+t).\end{aligned}$$

The set $\{-(i+3t): t \in \mathbf{R}\}$ is a straight line through $f(i)$ and Theorem 8 tells us that it is a good approximation for $f(i+t)$ near $f(i)$. It is in fact the tangent to $f(\gamma)$ at $f(i)$. It points in the direction determined by $\operatorname{Arg} f'(i) = \operatorname{Arg}(3i^2) = \pi$; in other words, it is also parallel to the real axis.

We could also look at the line through i parallel to the imaginary axis. Let us call it γ_1. It is the set $\{i(1+t): t \in \mathbf{R}\}$. The tangent to $f(\gamma_1)$ at $f(i)$ is

$$\{f(i) + f'(i) \cdot it : t \in \mathbf{R}\} = \{-i - 3it : t \in \mathbf{R}\}.$$

It points in the direction determined by $\operatorname{Arg} if'(i) = \operatorname{Arg}(-3i) = \frac{-\pi}{2}$. But $\operatorname{Arg} if'(i) - \operatorname{Arg} f'(i) = -\frac{3\pi}{2}$, so that $f(\gamma)$ and $f(\gamma_1)$ meet at right-angles as expected.

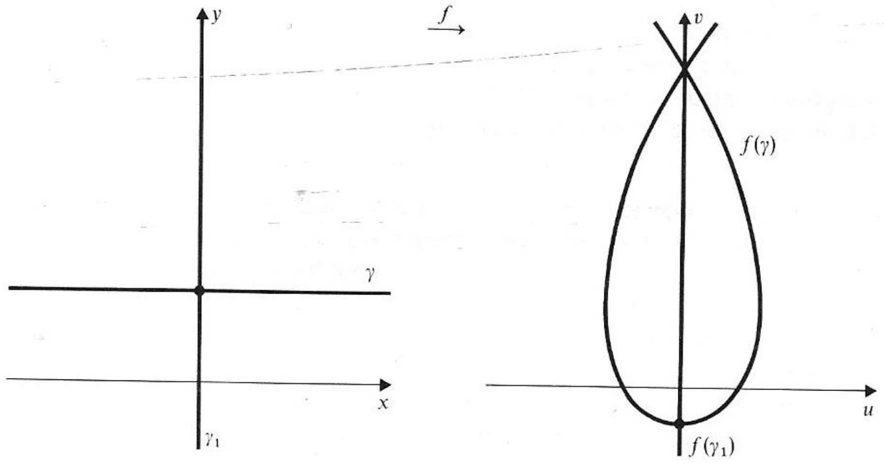


Fig. 5

As we pointed out in Section 2.1 of *Unit 2, Continuous Functions*, we may obtain some pictorial information about a function by examining how it transforms various curves. An obvious class of curves to look at is the net of *coordinate lines*—the lines parallel to the real, or imaginary, axes. Their images will form a net of curves, which divides into two families in such a way that each curve of either family intersects every curve of the other family at right angles. One family consists of the images of the lines parallel to the real axis, the other consists of images of lines parallel to the imaginary axis. Things may go wrong where the derivative of the function vanishes, though.

Example

We have already calculated the images of the coordinate lines parallel to the imaginary axis under the function $z \longrightarrow z^2$ in Section 2.1, *Unit 2*. We found that the image of the line $\operatorname{Re} z = \alpha$, $\alpha \neq 0$, parallel to the imaginary axis, is the parabola (in the w-plane) $u = \alpha^2 - \dfrac{v^2}{4\alpha^2}$ (Fig. 6). The image of the imaginary axis is the non-positive real axis.

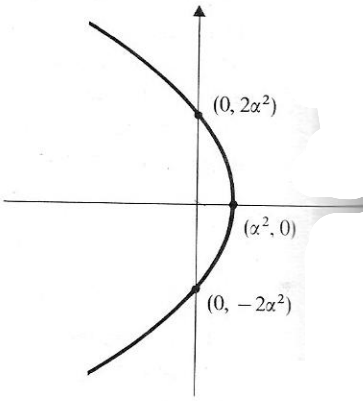


Fig. 6

The image of the line $\operatorname{Im} z = \beta$, $\beta \neq 0$, parallel to the real axis, is the parabola (in the w-plane) $u = \dfrac{v^2}{4\beta^2} - \beta^2$ (Fig. 7). The image of the real axis itself is the non-negative real axis.

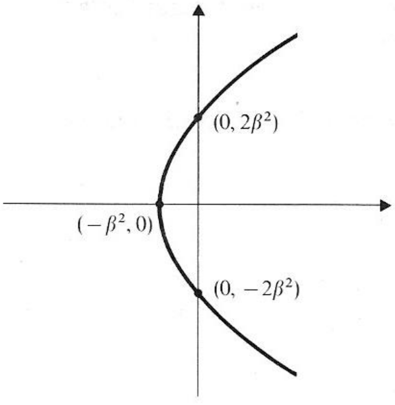


Fig. 7

Here is a picture of the images of several coordinate lines superimposed:

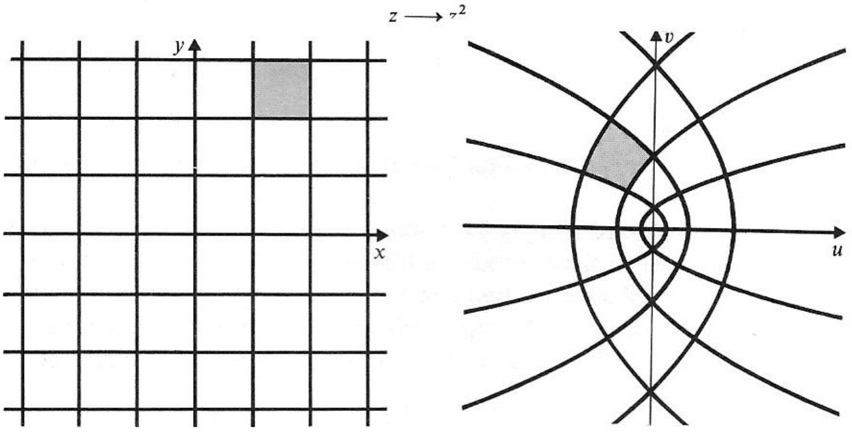


Fig. 8

Each parabola open towards the right is the image of a line parallel to the real axis, and so it intersects each parabola open to the left (which is the image of a line parallel to the imaginary axis) at right angles. We have chosen equally spaced lines in the z-plane. The image of a square in the z-plane is a lozenge shaped set in the w-plane: notice the variation in the sizes of the lozenges. This is another example of the magnification effect. Finally, notice the anomalous behaviour of the images of the real and imaginary axes at 0: the function $z \longrightarrow z^2$ has zero derivative at the origin.

Summary

In this section we have described how the derivative of a function at a point a may be interpreted geometrically, as a scale factor which gives the size and orientation of the image of a small disc centre a, and, more precisely, as a transformation of the tangents to curves through a.

Self-Assessment Questions

1. Show that the images of the imaginary and real axes under $z \longrightarrow z^3 + z$ intersect at right-angles.

2. By what factor does the function $z \longrightarrow z^2$ approximately magnify small discs with centre $-2\sqrt{3} - 2i$? Through what angle does it approximately rotate them?

3. What is the effect of the function $z \longrightarrow z + c$, where c is a fixed complex number, on the net of coordinate lines?

Solutions

1. The result follows since $f : z \longrightarrow z^3 + z$ is analytic on **C** and $f'(0) \neq 0$: hence the angle of intersection is preserved.

2. Magnification factor: 8; rotation angle: $-\dfrac{5\pi}{6}$.

3. It maps the net, as a whole, to itself, though individual coordinate lines are translated parallel to themselves.

3.3 PROBLEMS

1. Complete the proof of Theorem 3 of Section 3.1. Show that

   (i) if f and g are differentiable at a, then so is $f + g$, and
   $$(f + g)'(a) = f'(a) + g'(a);$$

   (ii) if f and g are differentiable at a and $g(a) \neq 0$, then f/g is differentiable at a, and
   $$(f/g)'(a) = \frac{f'(a) \cdot g(a) - f(a) \cdot g'(a)}{[g(a)]^2}.$$

2. (i) Let $p(z) = az^2 + bz + c$, where $a, b, c \in \mathbf{C}$. Show that $q(z) = \overline{p(\bar{z})}$ is given by $\bar{a}z^2 + \bar{b}z + \bar{c}$. Then show that the derivative of q at $\bar{\alpha}$ is $q'(\bar{\alpha}) = \overline{p'(\alpha)}$.

   (ii) The function f is differentiable at α. Define the function g by $g(z) = \overline{f(\bar{z})}$. Show that g is differentiable at $\bar{\alpha}$ and that $g'(\bar{\alpha}) = \overline{f'(\alpha)}$.

3. Show that if $p(z) = a_0 + a_1(z - a) + a_2(z - a)^2 + \cdots + a_n(z - a)^n$, then $a_k = \frac{1}{k!} p^{(k)}(a)$. (Here $p^{(k)}(a)$ is the kth derivative of p evaluated at a, if $k \geqslant 1$; $p^{(0)}(a) = p(a)$, and $0! = 1$.) Hence express $1 + z^2 + z^4$ as a polynomial in $(z - i)$.

4. Find the images of the coordinate lines under the function $z \longrightarrow \frac{1}{z}$ $(z \neq 0)$.

5. The ideas of the first part of Section 3.2 suggest another way of thinking about the Chain Rule. One would expect the scale factor of $f \circ g$ at a to be the product of the scale factor of g at a and the scale factor of f at $g(a)$. This gives the right formula for the derivative of $f \circ g$ at a. But can you prove the Chain Rule, using Theorem 8 (of Section 3.2)?

6. If γ_1 and γ_2 are two curves which intersect at a point a in a region on which a function f is analytic, and $|f|$ is constant on $\gamma_1 \cup \gamma_2$, what can you say about the curves and the function at a? (Hint: Try drawing the sets $f(\gamma_1)$ and $f(\gamma_2)$).

Solutions

1. (i) $\frac{(f + g)(a + h) - (f + g)(a)}{h} = \frac{f(a + h) - f(a)}{h} + \frac{g(a + h) - g(a)}{h}$, from which the differentiability of $f + g$ is immediately clear, and $(f + g)'(a) = f'(a) + g'(a)$.

   (ii) To establish the differentiability of f/g, we first prove that $1/g$ is differentiable.
   $$\left(\frac{1}{g(a + h)} - \frac{1}{g(a)}\right)\Big/h = -\frac{g(a + h) - g(a)}{h \cdot g(a + h) \cdot g(a)}$$
   which tends to $-\frac{g'(a)}{[g(a)]^2}$ as h approaches 0, using the continuity of g. We then use the fact that the product of differentiable functions is differentiable to establish that f/g is differentiable. The product rule gives:
   $$\left(f \cdot \frac{1}{g}\right)(a) = f'(a) \cdot \frac{1}{g(a)} + f(a) \cdot \left(-\frac{g'(a)}{[g(a)]^2}\right) = \frac{f'(a) \cdot g(a) - f(a) \cdot g'(a)}{[g(a)]^2}.$$

2. (i) Since $p(z) = az^2 + bz + c$, we have $p(\bar{z}) = a\bar{z}^2 + b\bar{z} + c$. Hence
   $$\overline{p(\bar{z})} = \overline{a\bar{z}^2 + b\bar{z} + c} = \bar{a}z^2 + \bar{b}z + \bar{c}.$$
   Now $q(z) = \overline{p(\bar{z})} = \bar{a}z^2 + \bar{b}z + \bar{c}$; hence $q'(z) = \bar{a}(2z) + \bar{b}(1)$, and so
   $$q'(\bar{\alpha}) = 2\bar{a}\bar{\alpha} + \bar{b}.$$

But $p'(z) = 2az + b$; hence $p'(\alpha) = 2a\alpha + b$, and so

$$\overline{p'(\alpha)} = 2\bar{a}\bar{\alpha} + \bar{b}.$$

Hence $q'(\bar{\alpha}) = \overline{p'(\alpha)}$.

(ii) $$\left|\frac{g(z) - g(\bar{\alpha})}{z - \bar{\alpha}} - \overline{f'(\alpha)}\right| = \left|\text{conjugate of } \frac{f(\bar{z}) - f(\alpha)}{\bar{z} - \alpha} - f'(\alpha)\right|$$
$$= \left|\frac{f(\bar{z}) - f(\alpha)}{\bar{z} - \alpha} - f'(\alpha)\right|$$

(remember that $|\bar{\zeta}| = |\zeta|$ for any complex number ζ).

It follows that $\lim_{z \to \bar{\alpha}} \frac{g(z) - g(\bar{\alpha})}{z - \bar{\alpha}} = \overline{f'(\alpha)}$. Here are the details. Let $\varepsilon > 0$, and let $\delta > 0$ be such that

$$\left|\frac{f(z) - f(\alpha)}{z - \alpha} - f'(\alpha)\right| < \varepsilon \text{ whenever } 0 < |z - \alpha| < \delta.$$

If $0 < |z - \bar{\alpha}| < \delta$, then $0 < |\bar{z} - \alpha| < \delta$ also, so $\left|\frac{f(\bar{z}) - f(\alpha)}{\bar{z} - \alpha} - f'(\alpha)\right| < \varepsilon$; in other words, $\left|\frac{g(z) - g(\bar{\alpha})}{z - \bar{\alpha}} - \overline{f'(\alpha)}\right| < \varepsilon$.

3. Differentiating $p(z)$ k times gives

$$p^{(k)}(z) = k!a_k + (k+1)\cdot k\cdot\ldots\cdot 2\cdot a_{k+1}\cdot(z - a) + \ldots$$
$$+ n\cdot(n-1)\cdot\ldots\cdot(n-k+1)\cdot a_n\cdot(z - a)^{n-k}.$$

Thus $a_k = \frac{1}{k!}p^{(k)}(a)$.

Suppose $1 + z^2 + z^4 = p(z) = a_0 + a_1(z - i) + a_2(z - i)^2 + \ldots + a_n(z - i)^n$. Now

$p(i) = 1$;

$p^{(1)}(z) = 2z + 4z^3$, so $p^{(1)}(i) = -2i$;

$p^{(2)}(z) = 2 + 12z^2$, so $p^{(2)}(i) = -10$;

$p^{(3)}(z) = 24z$, so $p^{(3)}(i) = 24i$;

$p^{(4)}(z) = 24$, so $p^{(4)}(i) = 24$.

Since $p^{(k)}(z) = 0$ for $k \geqslant 5$, $a_k = 0$ for $k \geqslant 5$, and

$$1 + z^2 + z^4 = 1 - 2i(z - i) - \frac{10}{2!}(z - i)^2 + \frac{24i}{3!}(z - i)^3 + \frac{24}{4!}(z - i)^4$$
$$= 1 - 2i(z - i) - 5(z - i)^2 + 4i(z - i)^3 + (z - i)^4.$$

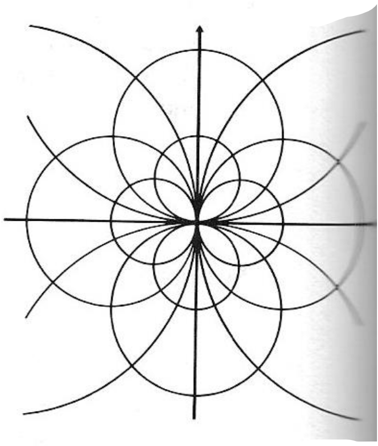


Fig. 9

4. A point on the line $\text{Im}\, z = \beta$ can be written $t + i\beta$, $t \in \mathbf{R}$. The image of the line $\text{Im}\, z = \beta$ is the set of points $\frac{1}{t + i\beta}$ or $\frac{t - i\beta}{t^2 + \beta^2}$. This is the set of points $u + iv$ (in the w-plane) such that $v = -\beta(u^2 + v^2)$. If $\beta \neq 0$ this is a circle whose centre is at $-\frac{i}{2\beta}$ and whose radius is $\frac{1}{2|\beta|}$. The image of the real axis (minus the origin) is again the real axis (minus the origin). The images of lines parallel to the imaginary axis are circles with centres on the real axis, except for the imaginary axis itself, whose image is the imaginary axis again.

5. We assume that g is differentiable at a, and that f is differentiable at $g(a)$; we wish to prove that $f \circ g$ is differentiable at a, and find its derivative. Since g is differentiable at a, there is a function ϕ such that

$$g(z) = g(a) + g'(a)\cdot(z - a) + \phi(z)$$

and $\lim_{z \to a} \frac{\phi(z)}{z - a} = 0$. Since f is differentiable at $g(a)$ there is a function ψ such that

$$f(z) = f(g(a)) + f'(g(a))\cdot(z - g(a)) + \psi(z)$$

and $\lim_{z \to g(a)} \frac{\psi(z)}{z - g(a)} = 0$. Thus

$$f(g(z)) = f(g(a)) + f'(g(a))\cdot(g(z) - g(a)) + \psi(g(z))$$
$$= f(g(a)) + f'(g(a))[g'(a)\cdot(z - a) + \phi(z)] + \psi(g(z))$$
$$= f(g(a)) + f'(g(a))\cdot g'(a)\cdot(z - a) + [f'(g(a))\cdot\phi(z) + \psi(g(z))].$$

The first two terms are just what we want; the theorem will be proved if we can show that $\lim_{z\to a}\dfrac{f'(g(a))\cdot\phi(z)+\psi(g(z))}{z-a}=0$.

It is certainly true that $\lim_{z\to a}\dfrac{f'(g(a))\cdot\phi(z)}{z-a}=0$, so the problem is to show that $\lim_{z\to a}\dfrac{\psi(g(z))}{z-a}=0$. Now provided $g(z)\neq g(a)$,

$$\frac{\psi(g(z))}{z-a}=\frac{\psi(g(z))}{g(z)-g(a)}\cdot\frac{g(z)-g(a)}{z-a}.$$

As z approaches a, $g(z)$ approaches $g(a)$ (since g is continuous at a). So $\dfrac{\psi(g(z))}{g(z)-g(a)}$ tends to zero (since $\lim_{z\to g(a)}\dfrac{\psi(z)}{z-g(a)}=0$). Moreover, $\lim_{z\to a}\dfrac{g(z)-g(a)}{z-a}=g'(a)$: so it would appear to follow that $\lim_{z\to a}\dfrac{\psi(g(z))}{z-a}=0$. Unfortunately, we have the usual problem: for the argument we have just given to apply, there must be a disc centre at a which contains no point z, other than a, such that $g(z)=g(a)$; and there may not be such a disc. We get over the problem in the usual way. Define a new function, χ, by

$$\chi(z)=\begin{cases}\dfrac{\psi(g(z))}{g(z)-g(a)}, & \text{if } g(z)\neq g(a)\\[2ex] 0 & \text{, otherwise.}\end{cases}$$

Then χ is continuous at a, $\chi(a)=0$, and

$$\frac{\psi(g(z))}{z-a}=\chi(z)\cdot\frac{g(z)-g(a)}{z-a}$$

(since $\psi(g(a))=0$, both sides are zero if $g(z)=g(a)$).

Thus $\lim_{z\to a}\dfrac{\psi(g(z))}{z-a}=\chi(a)\cdot g'(a)=0$, and the theorem is proved.

6. We know that $|f(z)|=k$, say, if z lies on γ_1 or γ_2: so the image sets $f(\gamma_1)$ and $f(\gamma_2)$ both lie on the circle $\{w:|w|=k\}$. The image curves are therefore identical, and so the angle between them at $f(a)$ is either zero or π. There are therefore only two possibilities: either γ_1 and γ_2 have a common tangent line at a, so that the angle between them at a is also zero or π, or on the other hand $f'(a)=0$.

3.4 THE EXPONENTIAL, TRIGONOMETRIC AND HYPERBOLIC FUNCTIONS

In this section we shall define complex functions which have most of the properties of those familiar real functions exp, cos and sin, and cosh and sinh, and which agree with those functions on the real axis. The similarities are sufficiently striking that we call our new complex functions by the same names.

* * * * * * * *

The Exponential Function

The real exponential function has two principal properties: $\exp(x_1 + x_2) = \exp x_1 \cdot \exp x_2$, and $\exp' = \exp$. We wish to find a complex function f which has analogous properties: $f(z_1 + z_2) = f(z_1) \cdot f(z_2)$, and $f' = f$. We also require that if z is real, $f(z) = \exp z = e^z$. We start with the first property, $f(z_1 + z_2) = f(z_1) \cdot f(z_2)$: can we find a function for which this is true? If it is to be true, then, in particular, it must be true that

$$f(x + iy) = f(x) \cdot f(iy) = (\exp x) \cdot f(iy).$$

Thus f is determined if we can find what it does to imaginary numbers. Let us put $f(iy) = u(y) + iv(y)$, where u and v are real functions. It must be true that

$$f(i(y_1 + y_2)) = f(iy_1) \cdot f(iy_2),$$

so

$$\begin{aligned} u(y_1 + y_2) + iv(y_1 + y_2) &= [u(y_1) + iv(y_1)] \cdot [u(y_2) + iv(y_2)] \\ &= [u(y_1) \cdot u(y_2) - v(y_1) \cdot v(y_2)] + i[u(y_1) \cdot v(y_2) + v(y_1) \cdot u(y_2)]. \end{aligned}$$

Equating real and imaginary parts gives the following addition formulas for u and v:

$$\begin{aligned} u(y_1 + y_2) &= u(y_1) \cdot u(y_2) - v(y_1) \cdot v(y_2) \\ v(y_1 + y_2) &= u(y_1) \cdot v(y_2) + v(y_1) \cdot u(y_2). \end{aligned}$$

But these have just the form of the addition formulas for the (real) functions cos and sin. If we set $u = \cos$ and $v = \sin$ these equations are satisfied, so we have a very likely candidate for the function $f: f(z) = e^x(\cos y + i \sin y)$, where $z = x + iy$. Does this function have the property that $f(z_1 + z_2) = f(z_1) \cdot f(z_2)$? (Our argument so far does not show that this is true for *all* z_1 and z_2.) The following calculation clears that point up.

$$\begin{aligned} f(z_1 + z_2) &= f((x_1 + x_2) + i(y_1 + y_2)) \\ &= \exp(x_1 + x_2) \cdot f(i(y_1 + y_2)) \\ &= (\exp x_1) \cdot (\exp x_2) \cdot f(iy_1) \cdot f(iy_2) \\ &= f(z_1) \cdot f(z_2). \end{aligned}$$

The other question we need to answer is this: is f differentiable? If so, does it satisfy $f' = f$? Well,

$$\begin{aligned} \frac{f(z + h) - f(z)}{h} &= \frac{f(z) \cdot f(h) - f(z)}{h} \\ &= f(z) \cdot \frac{f(h) - f(0)}{h}, \quad \text{since } f(0) = 1. \end{aligned}$$

Thus f is differentiable at z if and only if it is differentiable at 0. To get some idea of what the derivative of f at 0 might be, we consider the first few terms in the Taylor series at 0 of e^x, $\cos y$ and $\sin y$. Now

$$\begin{aligned} e^x &= 1 + x + \theta(x) \\ \cos y &= 1 + \phi(y) \\ \sin y &= y + \psi(y), \end{aligned}$$

where θ, ϕ and ψ are remainder terms. Thus

$$\begin{aligned} f(z) = e^x(\cos y + i \sin y) &= [1 + x + \theta(x)][1 + \phi(y) + iy + i\psi(y)] \\ &= 1 + x + iy + \Phi(x, y), \end{aligned}$$

where Φ is a collection of remainder terms. So we may write

$$f(z) = f(0) + z + \Phi(z),$$

which suggests that f is differentiable at 0 with derivative 1. To establish that this really is so, we must examine the remainder term more closely, and apply Theorem 8 (of Section 3.2). This we do in the following theorem.

Theorem 9

Let $f(z) = e^x(\cos y + i \sin y)$, where $z = x + iy$. Then f is differentiable at 0, and $f'(0) = 1$. It follows that f is entire, and that $f' = f$.

Proof

Let $\Phi(z) = f(z) - f(0) - z$. Then

$$\begin{aligned} \Phi(z) &= e^x(\cos y + i \sin y) - 1 - (x + iy) \\ &= (e^x \cos y - 1 - x) + i(e^x \sin y \underset{-}{\cancel{+}} y). \end{aligned}$$

In the next step, we rearrange the expression for $\Phi(z)$ so that we may use the differentiability of the (real) functions $x \longrightarrow e^x$, cos and sin, at 0. (Remember that the derivatives of these functions at 0 are 1, 0 and 1 respectively.)

$$\begin{aligned} \Phi(z) = {} & (e^x - 1 - x)\cos y + (1 + x)(\cos y - 1) \\ & + i[e^x(\sin y - y) + (e^x - 1)y]. \end{aligned} \qquad (*)$$

We have to show that $\lim_{z \to 0} \frac{\Phi(z)}{z} = 0$. Now if x and y are both non-zero,

$$\begin{aligned} \frac{\Phi(z)}{z} = {} & \left(\frac{e^x - 1}{x} - 1\right) \cdot \cos y \cdot \frac{x}{z} + (1 + x) \cdot \left(\frac{\cos y - 1}{y}\right) \cdot \frac{y}{z} \\ & + i\left[e^x \cdot \left(\frac{\sin y}{y} - 1\right) \cdot \frac{y}{z} + (e^x - 1) \cdot \frac{y}{z}\right]. \end{aligned}$$

But $|x| = |\text{Re}\, z| \leqslant |z|$, and $|y| = |\text{Im}\, z| \leqslant z$, so $\left|\frac{x}{z}\right| \leqslant 1$ and $\left|\frac{y}{z}\right| \leqslant 1$. Using these inequalities, and the familiar fact that $|\cos y| \leqslant 1$, we obtain

$$\begin{aligned} \left|\frac{\Phi(z)}{z}\right| \leqslant {} & \left|\frac{e^x - 1}{x} - 1\right| + |1 + x| \cdot \left|\frac{\cos y - 1}{y}\right| \\ & + |e^x| \cdot \left|\frac{\sin y}{y} - 1\right| + |e^x - 1|. \end{aligned}$$

If $y = 0$, but $x \neq 0$, then from $(*)$

$$\left|\frac{\Phi(z)}{z}\right| = \left|\frac{e^x - 1}{x} - 1\right|,$$

while if $x = 0$, but $y \neq 0$, then from $(*)$

$$\left|\frac{\Phi(z)}{z}\right| = \left|\frac{\cos y - 1}{y} + i\left(\frac{\sin y}{y} - 1\right)\right|$$

$$\leqslant \left|\frac{\cos y - 1}{y}\right| + \left|\frac{\sin y}{y} - 1\right|.$$

We know that $\lim_{x \to 0}\left(\frac{e^x - 1}{x} - 1\right) = 0$, $\lim_{y \to 0}\left(\frac{\cos y - 1}{y}\right) = 0$, and $\lim_{y \to 0}\left(\frac{\sin y}{y} - 1\right) = 0$; since the function $x \longrightarrow e^x$ is continuous at 0, $\lim_{x \to 0}(e^x - 1) = 0$. Moreover, as z approaches 0, so x and y both approach 0. We conclude that $\lim_{z \to 0} \frac{\Phi(z)}{z} = 0$, and so by Theorem 8, f is differentiable at 0 and $f'(0) = 1$. Since

$$\frac{f(z+h) - f(z)}{h} = f(z) \cdot \left(\frac{f(h) - f(0)}{h}\right),$$

f is differentiable at z and $f'(z) = f(z) \cdot f'(0) = f(z)$. ■

(In Section 3.8 we give a shorter proof of this result: this shorter proof, in terms of the Cauchy-Riemann equations, demands a knowledge of partial differentiation.)

Since the function $f(z) = e^x(\cos y + i \sin y)$ has the most important properties of the real exponential function, we call it the (complex) exponential function.

Definition

The **exponential function**, exp, is the function defined by

$$\exp z = e^x(\cos y + i \sin y) \quad \text{for all } z \in \mathbf{C},$$

where $z = x + iy$.

We shall often use the notation e^z for $\exp z$. Broadly speaking, we shall use exp for the name of the function, and e^z for its value, except where difficulties of printing make exp preferable. This notation has its disadvantages—it is rather tempting to read e^z as "e to the power z", which does not make much sense—but it is pretty well universal.

Properties of exp

(i) $e^0 = 1$.

(ii) $|e^z| = e^x$
(since $|\cos y + i \sin y| = 1$).

(iii) e^z is not zero for any z
(since $|e^z| = e^x$, and e^x can never be zero).

(iv) $\frac{1}{e^z} = e^{-z}$
(since $e^z \cdot e^{-z} = e^0 = 1$).

(v) $(e^z)^n = e^{nz}$ for any integer n.
(In particular $(e^z)^2 = e^z \cdot e^z = e^{2z}$; the general result is proved by induction.)

(vi) $e^{i\pi} = -1$, and $e^{2i\pi} = 1$.

There is a very significant difference between the complex and the real exponential functions. The real exponential function is one-one; but since $e^{2i\pi} = 1 = e^0$, the complex exponential function certainly is not one-one. In fact, it is periodic.

Definition

A complex function f is **periodic** if there is a non-zero complex number p such that $f(z + p) = f(z)$ for all z. The number p is said to be a **period** of f.

(A periodic function has many periods: it is easy to see that if p is a period, then so is np where n is any non-zero integer. The usual situation is that all periods arise in this way, as multiples of one basic period.) We can now add another to our list of properties.

(vii) exp is periodic, and its periods are the non-zero integer multiples of $2i\pi$.

(To see that $2i\pi$ is a period, notice that

$$e^{z+2i\pi} = e^z \cdot e^{2i\pi} = e^z.$$

On the other hand, suppose that p is a period. Then $e^{z+p} = e^z$ for all z, so in particular $e^p = e^{0+p} = e^0$. Thus $e^p = 1$. Suppose that $p = \alpha + i\beta$. Then from property (ii) $|e^p| = e^\alpha = 1$, and so $\alpha = 0$; and $\cos\beta + i\sin\beta = 1$, so $\beta = 2n\pi$ for some integer n. If p is a period it must be non-zero; so finally $p = 2ni\pi$ where n is a non-zero integer.

Notice that during the course of this argument we have shown that if $e^p = 1$, then $p = 2ni\pi$ for some integer n, which is a very useful fact.)

To get some kind of pictorial idea of the complex exponential function, we can look at its effect on coordinate lines. A point on the line $\operatorname{Im} z = \beta$, which is parallel to the real axis, can be written $t + i\beta$, $t \in \mathbf{R}$. This point is mapped to $e^t(\cos\beta + i\sin\beta)$ and so the line is mapped to the curve $\{e^t(\cos\beta + i\sin\beta) : t \in \mathbf{R}\}$. This is a radial half-line or *ray* (not containing the origin).

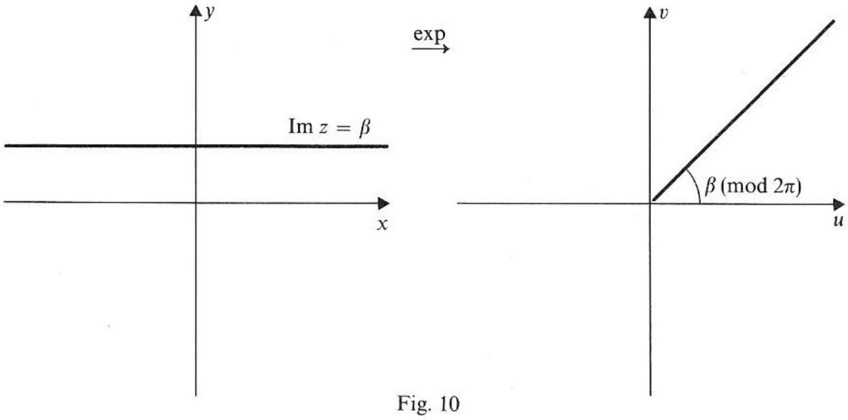


Fig. 10

The line $\operatorname{Re} z = \alpha$, that is $\{\alpha + it : t \in \mathbf{R},\ \alpha \in \mathbf{R}\}$, parallel to the imaginary axis, is mapped to $\{e^\alpha(\cos t + i\sin t) : t \in \mathbf{R}\}$, which is a circle of radius e^α.

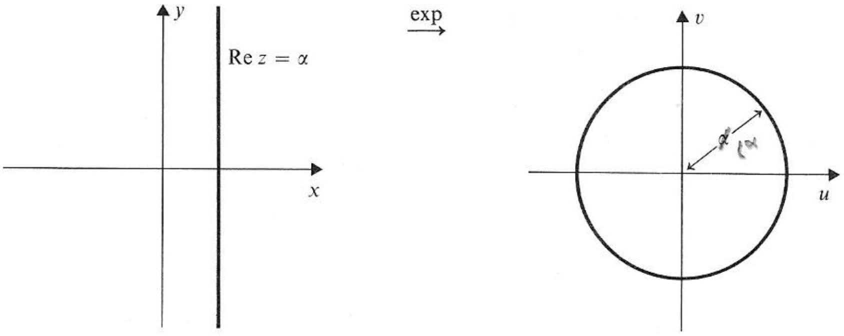


Fig. 11

So the net of coordinate lines is mapped to a net of concentric circles and radial half-lines.

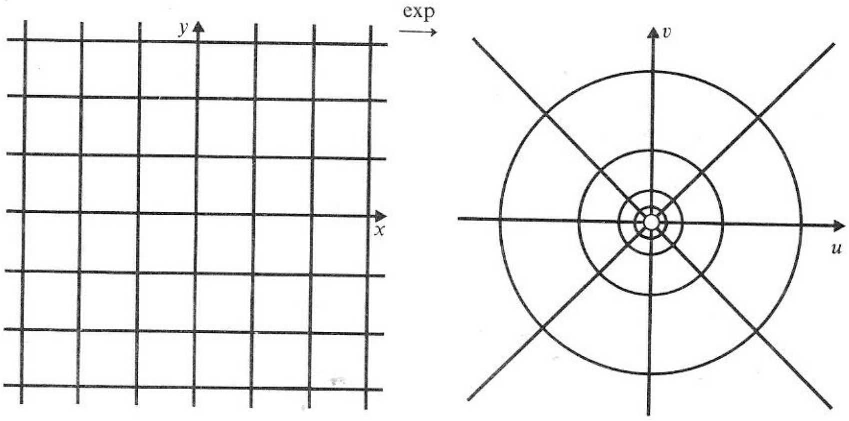


Fig. 12

Self-Assessment Questions

1. Express e^{-z} in terms of x and y (where $z = x + iy$).
2. Calculate the derivatives of the functions
   (i) $z \longrightarrow \exp(z^2)$;
   (ii) $z \longrightarrow (\exp z)^2$;
   (iii) $z \longrightarrow \exp(e^z)$.
3. Show that $\exp(\bar{z}) = \overline{\exp z}$ for all z.

Solutions

1. $e^{-z} = e^{-x-iy} = e^{-x}[\cos(-y) + i\sin(-y)]$
   $= e^{-x}(\cos y - i \sin y)$.
2. (i) $2z\exp(z^2)$. (ii) $2(\exp z)^2$. (iii) $e^z\exp(e^z)$.
3. $\exp(\bar{z}) = e^{x-iy} = e^x(\cos y - i\sin y) = \overline{\exp z}$.

The Trigonometric Functions

We now turn our attention to the problem of finding complex functions whose properties mimic the properties of the real trigonometric functions cos and sin.

Our definition of exp involves (real) cos and sin; in fact, we can express those functions in terms of exp. This suggests a way of extending the trigonometric functions to the complex plane which, as it turns out, is very satisfactory. From the definition of exp, if t is any real number,

$$e^{it} = \cos t + i \sin t \quad \text{and} \quad e^{-it} = \cos t - i \sin t.$$

Thus

$$\cos t = \frac{1}{2}(e^{it} + e^{-it}) \quad \text{and} \quad \sin t = \frac{1}{2i}(e^{it} - e^{-it}).$$

These formulas suggest that we define complex cos and sin functions as follows.

Definition

The (complex) trigonometric functions, cos and sin, are defined by

$$\cos z = \frac{1}{2}(e^{iz} + e^{-iz})$$

$$\sin z = \frac{1}{2i}(e^{iz} - e^{-iz}) \qquad \text{for all } z \in \mathbf{C}.$$

Properties of cos and sin

(i) cos and sin are entire functions; $\cos' = -\sin$, $\sin' = \cos$.

$$\left(\cos' z = \frac{1}{2}(ie^{iz} - ie^{-iz}) = -\frac{1}{2i}(e^{iz} - e^{-iz}) = -\sin z;\right.$$

$$\left.\sin' z = \frac{1}{2i}(ie^{iz} + ie^{-iz}) = \frac{1}{2}(e^{iz} + e^{-iz}) = \cos z.\right)$$

(ii) cos is an even function—that is, $\cos(-z) = \cos z$: and sin is odd—that is, $\sin(-z) = -\sin z$.

(See Self-Assessment Question 4.)

(iii) $\cos^2 z + \sin^2 z = 1$.

$$\left(\cos^2 z + \sin^2 z = \frac{1}{4}(e^{2iz} + 2 + e^{-2iz}) - \frac{1}{4}(e^{2iz} - 2 + e^{-2iz})\right.$$
$$\left.= 1.\right)$$

(iv) $\cos(z + w) = \cos z \cos w - \sin z \sin w$, and

$\sin(z + w) = \sin z \cos w + \cos z \sin w$.

(We prove the addition formula for cos, and ask you to prove the one for sin in Self-Assessment Question 5.

$$\begin{aligned}\cos z \cos w - \sin z \sin w &= \frac{1}{4}(e^{iz} + e^{-iz})(e^{iw} + e^{-iw}) \\ &\quad + \frac{1}{4}(e^{iz} - e^{-iz})(e^{iw} - e^{-iw}) \\ &= \frac{1}{4}(e^{i(z+w)} + e^{i(z-w)} + e^{-i(z-w)} + e^{-i(z+w)}) \\ &\quad + \frac{1}{4}(e^{i(z+w)} - e^{i(z-w)} - e^{-i(z-w)} + e^{-i(z+w)}) \\ &= \frac{1}{2}(e^{i(z+w)} + e^{-i(z+w)}) = \cos(z + w).)\end{aligned}$$

(v) $\cos z = \sin\left(\frac{\pi}{2} - z\right)$.

$$\left(\sin\left(\frac{\pi}{2} - z\right) = \frac{1}{2i}\left[\exp\left(i\left(\frac{\pi}{2} - z\right)\right) - \exp\left(-i\left(\frac{\pi}{2} - z\right)\right)\right]\right.$$
$$\left.= \frac{1}{2i}[ie^{-iz} - (-i)e^{iz}] = \cos z.\right)$$

(vi) $\cos z = 0$ if and only if $z = \frac{\pi}{2} + n\pi$, where n is an integer; $\sin z = 0$ if and only if $z = n\pi$, where n is an integer.

(We deal with sin, leaving cos for you to tackle in Self-Assessment Question 6.

We have $\sin z = 0$ if and only if $e^{iz} = e^{-iz}$, that is, if and only if $e^{2iz} = 1$. But, as mentioned on page 121, $e^{2iz} = 1$ if and only if $2iz = 2in\pi$ for some integer n. Thus $\sin z = 0$ if and only if $z = n\pi$, where n is an integer.)

(vii) Both cos and sin are periodic, with period 2π.

(This follows immediately from the fact that $2i\pi$ is a period of exp.)

You may have noticed a certain resemblance between the definitions of the complex trigonometric functions and the real hyperbolic functions. The next two items on our list examine the effects of this resemblance, and show that the relation between the trigonometric functions and the hyperbolic functions is really very close. One expects there to be such a relationship because of the obvious similarities between such formulas as $\cos^2 x + \sin^2 x = 1$ and $\cosh^2 x - \sinh^2 x = 1$; but it is never satisfactorily explained in real analysis.

(viii) If t is a real number, then $\cos(it) = \cosh t$ and $\sin(it) = i \sinh t$.

(In fact $\cos(it) = \frac{1}{2}(e^{-t} + e^{t}) = \cosh t$; the other formula is proved similarly.

Notice that if z is not a real number then it is not necessarily true that $|\cos z| \leqslant 1$ or $|\sin z| \leqslant 1$.)

(ix) $\cos(x + iy) = \cos x \cosh y - i \sin x \sinh y$

$\sin(x + iy) = \sin x \cosh y + i \cos x \sinh y$.

(These formulas follow then from the addition formulas, (iv), and the formulas of (viii).)

Self-Assessment Questions

4. Show that cos is an even function and that sin is an odd function.
5. Show that $\sin(z + w) = \sin z \cos w + \cos z \sin w$.
6. Show that $\cos z = 0$ if and only if $z = \frac{\pi}{2} + n\pi$, where n is an integer.

Solutions

4. $\cos(-z) = \frac{1}{2}(e^{-iz} + e^{iz}) = \cos z$.

$\sin(-z) = \frac{1}{2i}(e^{-iz} - e^{iz}) = -\sin z$.

5. $\sin z \cos w + \cos z \sin w$

$$= \frac{1}{4i}(e^{iz} - e^{-iz})(e^{iw} + e^{-iw})$$

$$+ \frac{1}{4i}(e^{iz} + e^{-iz})(e^{iw} - e^{-iw})$$

$$= \frac{1}{4i}(e^{i(z+w)} + e^{i(z-w)} - e^{-i(z-w)} - e^{-i(z+w)})$$

$$+ \frac{1}{4i}(e^{i(z+w)} - e^{i(z-w)} + e^{-i(z-w)} - e^{-i(z+w)})$$

$$= \frac{1}{2i}(e^{i(z+w)} - e^{-i(z+w)}) = \sin(z + w).$$

6. We have shown that $\sin z = 0$ if and only if $z = n\pi$, where n is an integer.

Therefore since $\cos z = \sin\left(\frac{\pi}{2} - z\right)$, by property (v), $\cos z = 0$ if and only if $\frac{\pi}{2} - z = n\pi$, that is $z = \frac{\pi}{2} + n\pi$, where n is an integer.

The Hyperbolic Functions

The complex hyperbolic functions are defined in exactly the same way as their real counterparts.

Definition

The (complex) hyperbolic functions, cosh and sinh, are defined by

$$\cosh z = \tfrac{1}{2}(e^z + e^{-z})$$

$$\sinh z = \tfrac{1}{2}(e^z - e^{-z}), \quad \text{for all } z \in \mathbf{C}.$$

We could investigate the properties of these functions directly from the definition, but it is simpler to observe that

$$\cosh z = \cos(iz)$$

and

$$\sinh z = -i\sin(iz),$$

and deduce the properties of cosh and sinh from those of cos and sin. The following properties of the hyperbolic functions are more or less immediate.

Properties of cosh and sinh

(i) cosh and sinh are entire functions; $\cosh' = \sinh$, $\sinh' = \cosh$.

(ii) cosh is an even function, and sinh is odd.

(iii) $\cosh^2 z - \sinh^2 z = 1$.

(iv) $\cosh(z + w) = \cosh z \cosh w + \sinh z \sinh w$,
$\sinh(z + w) = \sinh z \cosh w + \cosh z \sinh w$.

(v) $\cosh z = 0$ if and only if $z = i\frac{\pi}{2} + n i\pi$, and $\sinh z = 0$ if and only if $z = n i\pi$, where n is an integer.

(vi) Both cosh and sinh are periodic, with period $2i\pi$.

Self-Assessment Questions

7. Just to make quite sure that you have understood, derive the formulas for (i) $\cosh'$ and (ii) $\cosh(z + w)$.

8. Establish formulas for the real and imaginary parts of (i) $\cosh z$ and (ii) $\sinh z$.

Solutions

7. (i) $\dfrac{d}{dz}\cosh z = \dfrac{d}{dz}\cos(iz) = -i\sin(iz) = \sinh z.$

(ii) $$\begin{aligned}\cosh(z+w) &= \cos(iz+iw)\\ &= \cos(iz)\cos(iw) - \sin(iz)\sin(iw)\\ &= \cos(iz)\cos(iw) + [-i\sin(iz)][-i\sin(iw)]\\ &= \cosh z\cosh w + \sinh z\sinh w.\end{aligned}$$

8. (i) $$\begin{aligned}\cosh(x+iy) &= \cosh x\cosh(iy) + \sinh x\sinh(iy)\\ &= \cosh x\cos y + i\sinh x\sin y.\end{aligned}$$

So the real part of $\cosh z$ is $\cosh x \cdot \cos y$, and the imaginary part is $\sinh x \cdot \sin y$.

(ii) $$\begin{aligned}\sinh(x+iy) &= \sinh x\cosh(iy) + \cosh x\sinh(iy)\\ &= \sinh x\cos y + i\cosh x\sin y.\end{aligned}$$

So the real part of $\sinh z$ is $\sinh x \cdot \cos y$, and the imaginary part is $\cosh x \cdot \sin y$.

Summary

We have found a complex function whose properties are very similar to those of the real exponential function, and which agrees with that function on the real axis. We accordingly named it the complex exponential function. We found one significant way in which it differs from the real exponential function: it is not one-one—in fact, it is periodic.

Using the complex exponential function we were able to define extensions of the real trigonometric and hyperbolic functions to the complex plane. We discovered that these complex trigonometric and hyperbolic functions are closely related to each other.

3.5 PROBLEMS

The first two problems are important—as you will see in the next section.

1. (i) Show that if w is any non-zero complex number, then there is some complex number z such that $w = e^z$.
   (Hint: Use polar coordinates.)

   (ii) If z_1 and z_2 satisfy $\exp z_1 = \exp z_2 = w$, what is the relationship between z_1 and z_2?

2. Find a set on which exp is one-one. Better still, find a *region* R such that exp is one-one on R, but not on any region properly containing R. (There are many different answers to this problem.)

3. Show that all periods of cos and sin are non-zero integer multiples of 2π.

4. Solve the equation $\cos z = 2$.
   (Hint: Use the formula $\cos(x + iy) = \cos x \cosh y - i \sin x \sinh y$.)

5. Show that $\cos \bar{z} = \overline{\cos z}$.

6. Find the images of the coordinate lines under the function $z \longrightarrow \sin z$.

7. We define the function tanh in the obvious way:

$$\tanh z = \frac{\sinh z}{\cosh z}.$$

   What is the domain of tanh? Where is it analytic? What is its derivative?

Solutions

1. (i) The number w can be written in polar form as $|w|(\cos\theta + i\sin\theta)$, where $|w| \neq 0$ since $w \neq 0$, and θ is an argument of w. Thus, if $e^z = w$, then

$$e^x(\cos y + i \sin y) = |w|(\cos\theta + i\sin\theta).$$

   It follows that $e^x = |w|$, and so $x = \log|w|$; and $y = \theta + 2k\pi$, where k is an integer. This condition on y just means that y has to be an argument of w. So

   if $z = \log|w| + i\theta$, where θ is any argument of ~~z~~ w, then $e^z = w$.

   (ii) If $\exp z_1 = \exp z_2 = w$, then $z_1 = \log|w| + i\theta_1$ and $z_2 = \log|w| + i\theta_2$, where θ_1 and θ_2 are both arguments of w. Thus z_1 and z_2 differ by an integer multiple of $2i\pi$.

2. The function exp is one-one on the real axis, for example: but we can do better than that. Suppose S is a set on which exp is one-one. Then we know that if $z \in S$, then none of the numbers $z + 2ni\pi$, where n is a non-zero integer, can be in S. This means that S can have maximum width, measured parallel to the imaginary axis, of 2π. An example of a *region* R whose width does not exceed 2π is $\{z: -\pi < \operatorname{Im} z < \pi\}$ (Fig. 13).

   Now exp is certainly one-one on R: but can we find any larger region R' (one containing R and something else beside) on which exp is one-one? Well, R' certainly cannot contain any point z for which $\operatorname{Im} z > \pi$ or $\operatorname{Im} z < -\pi$, since if it did it would contain points differing by $2i\pi$. Also we cannot obtain a *region* by adjoining to R points which satisfy $\operatorname{Im} z = \pi$ or $\operatorname{Im} z = -\pi$. So we conclude that exp is one-one on $\{z: -\pi < \operatorname{Im} z < \pi\}$ but not on any region properly containing it.

   There are many other solutions. One obvious one, which you may have thought of, is $\{z: 0 < \operatorname{Im} z < 2\pi\}$. Or you might even have thought of something like the one in Fig. 14. Our reasons for preferring $\{z: -\pi < \operatorname{Im} z < \pi\}$ will become clear later.

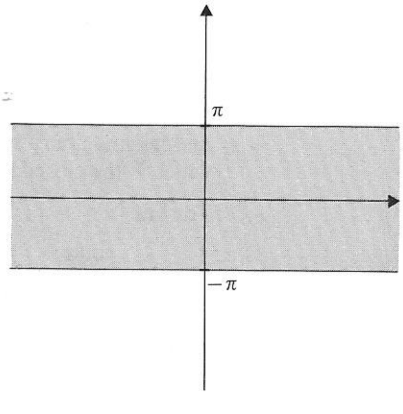


Fig. 13

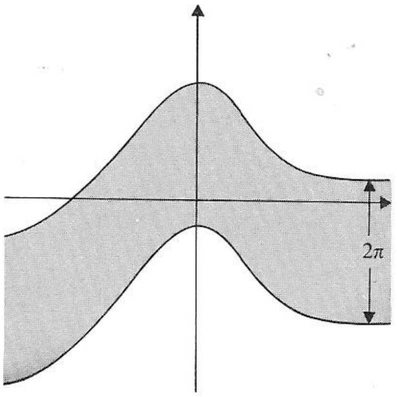


Fig. 14

3. If p is a period of cos, then $\cos(z + p) = \cos z$ for all z; in other words, $\cos z \cos p - \sin z \sin p = \cos z$ for all z. Thus $\sin p = 0$ (by putting $z = \pi/2$), and $\cos p = 1$. The solutions of $\sin p = 0$ are $p = n\pi$ (by property (vi), page 123) and $\cos n\pi = 1$ only if n is even. Thus all the periods of cos are of the form $2n\pi$, where n is a non-zero integer.

   By property (v), page 123, $\cos z = \sin(\pi/2 - z)$, so $\sin z = \cos(\pi/2 - z) = \cos(z - \pi/2)$, since cos is even. Hence since the periods of cos are of the form $2n\pi$, where n is a non-zero integer, so are those of sin.

4. If $\cos z = 2$, and $z = x + iy$, then
$$\cos x \cosh y - i \sin x \sinh y = 2.$$
   Equating real and imaginary parts, we have
$$\cos x \cosh y = 2, \quad \sin x \sinh y = 0.$$
   Thus $\sinh y = 0$ or $\sin x = 0$. The only solution to the equation $\sinh y = 0$ is $y = 0$; but then $\cosh y = 1$, and the equation $\cos z = 2$ has no solutions. So we must have $\sin x = 0$, that is $x = n\pi$, where n is an integer. But $\cosh y > 0$, so we must have $\cos x > 0$, which means that $x = n\pi$ with n an even integer, say $n = 2m$. This leaves us with $\cosh y = 2$ (since $\cos x = 1$). Now $\cosh y = \frac{1}{2}(e^y + e^{-y})$, so if $\cosh y = 2$ then $e^y + e^{-y} = 4$, or $e^{2y} - 4e^y + 1 = 0$. This quadratic equation for e^y has roots $(4 \pm \sqrt{16 - 4})/2 = 2 \pm \sqrt{3}$. Thus $e^y = 2 \pm \sqrt{3}$, and $y = \log(2 \pm \sqrt{3})$.

   So, finally, if $\cos z = 2$ then $z = 2m\pi + i\log(2 \pm \sqrt{3})$. It is easy to check that these complex numbers do satisfy the equation. Thus the solutions of the equation $\cos z = 2$ are the complex numbers $2m\pi + i\log(2 \pm \sqrt{3})$, where m is an integer.

5. $$\begin{aligned}\cos \bar{z} &= \tfrac{1}{2}[\exp(i\bar{z}) + \exp(-i\bar{z})]\\ &= \tfrac{1}{2}[\exp(\overline{-iz}) + \exp(\overline{iz})]\\ &= \tfrac{1}{2}[\overline{\exp(-iz)} + \overline{\exp(iz)}], \quad \text{since } \exp(\bar{z}) = \overline{\exp(z)},\\ &= \overline{\cos z}.\end{aligned}$$

6. Consider first the line parallel to the real axis, $\{t + i\beta : t \in \mathbf{R}\}$, β a constant real number. The image of this line is the set $\{\sin(t + i\beta) : t \in \mathbf{R}\}$. Now
$$\sin(t + i\beta) = \sin t \cosh \beta + i \cos t \sinh \beta = u + iv, \text{ say;}$$
   so the image is $\dfrac{u^2}{(\cosh \beta)^2} + \dfrac{v^2}{(\sinh \beta)^2} = 1$, provided $\beta \neq 0$ so that $\sinh \beta \neq 0$ ($\cosh \beta \neq 0$ for real β, of course). This is an ellipse which intersects the real axis at $u = \pm\cosh \beta$, and the imaginary axis at $v = \pm\sinh \beta$. The image of the real axis itself is $\{\sin t : t \in \mathbf{R}\}$, which is the closed interval of the real axis from -1 to 1.

   The line $\{\alpha + it : t \in \mathbf{R}\}$, α a constant real number, is mapped to $\{\sin(\alpha + it) : t \in \mathbf{R}\}$. Now
$$\sin(\alpha + it) = \sin \alpha \cosh t + i \cos \alpha \sinh t = u + iv, \quad \text{say.}$$
   Provided that α is not a multiple of $\pi/2$ (so that neither $\sin \alpha$ or $\cos \alpha$ is zero) the image is the hyperbola $\dfrac{u^2}{(\sin \alpha)^2} - \dfrac{v^2}{(\cos \alpha)^2} = 1$, which intersects the real axis at $u = \pm\sin \alpha$. If α is an even multiple of $\pi/2$, so that $\sin \alpha = 0$, the hyperbola degenerates to the imaginary axis.

   If α is an odd multiple of $\pi/2$, two cases arise. (i) If $\alpha = (4k + 1)\pi/2$, where k is an integer, the image is the interval $[1, \infty)$ of the real axis. (ii) If $\alpha = (4k + 3)\pi/2$, where k is an integer, the image is the interval $(-\infty, -1]$ of the real axis.

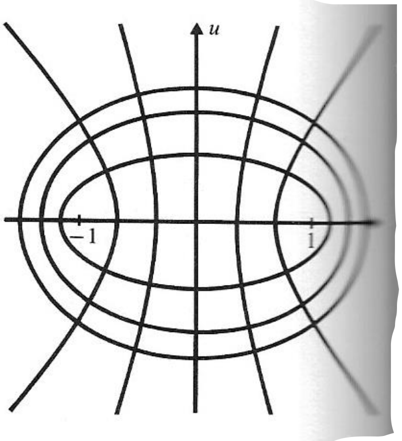


Fig. 15

7. The domain of tanh is $\{z : \cosh z \neq 0\}$, that is, all complex numbers except those of the form $i(\pi/2) + ni\pi$. It is analytic on its domain (which is a region). Its derivative is
$$\tanh' z = \frac{\cosh^2 z - \sinh^2 z}{\cosh^2 z} = \frac{1}{\cosh^2 z}.$$

3.6 LOGARITHMS AND POWERS

The logarithm function of real analysis is the inverse of the exponential function. In complex analysis the exponential function is not one-one, and so we cannot straightforwardly define a logarithm. We can define an inverse to the exponential function provided we restrict its domain so as to make it one-one.

We use a similar stratagem in real analysis in defining an inverse of the sine function, for example. The function sin is many-one on its domain, **R**, and so has no inverse function, properly speaking. The function $x \longrightarrow \sin x$, $x \in (-\pi/2, \pi/2)$, on the other hand, *is* one-one; and we can define its inverse function, arcsin. One way of specifying the function arcsin is by the rules

$$\sin(\arcsin x) = x \quad \text{for } x \in (-1, 1);$$

$$\arcsin x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right).$$

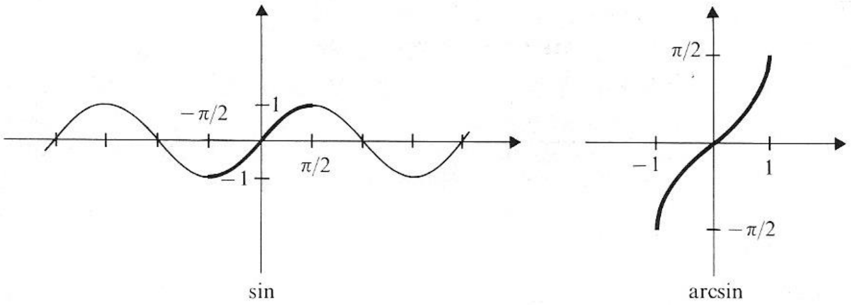


Fig. 16

There are other intervals on which sin is one-one: for example, the function $x \longrightarrow \sin x$, $x \in (\pi/2, 3\pi/2)$, is one-one. This function has an inverse, S, such that

$$\sin(S(x)) = x \quad \text{for } x \in (-1, 1);$$

$$S(x) \in \left(\frac{\pi}{2}, \frac{3\pi}{2}\right).$$

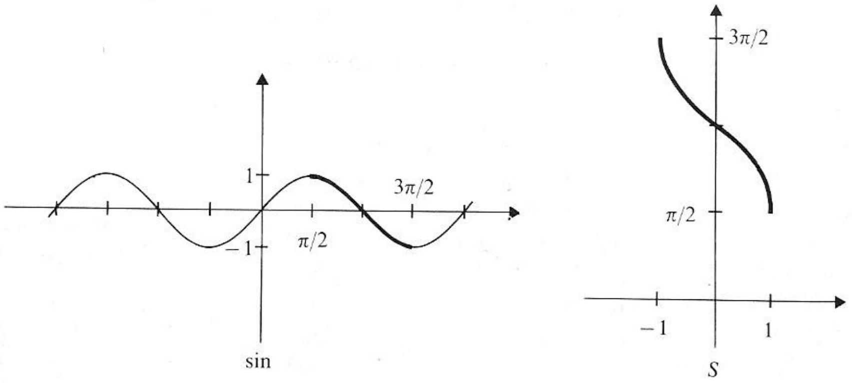


Fig. 17

It is usually enough to deal with arcsin, but you should be aware that there are other functions, like S, that can be described, loosely, as inverses of sin.

There are many subsets of the complex plane on which exp is one-one; for the moment we restrict our attention to one, the region $\{z : -\pi < \operatorname{Im} z < \pi\}$. (See Problem 2 of Section 3.5.) The exponential function maps this region one-one onto the region which consists of the whole complex plane with the exception of the nonpositive part of the real axis. For, if $w = \exp z = e^x(\cos y + i \sin y)$ then $|w|$ may take any positive value and Arg w (equal to y) lies between $-\pi$ and π. The exponential function on $\{z : -\pi < \operatorname{Im} z < \pi\}$ has an inverse, which maps the complex plane with the nonpositive real axis removed, back to the strip $-\pi < \operatorname{Im} z < \pi$.

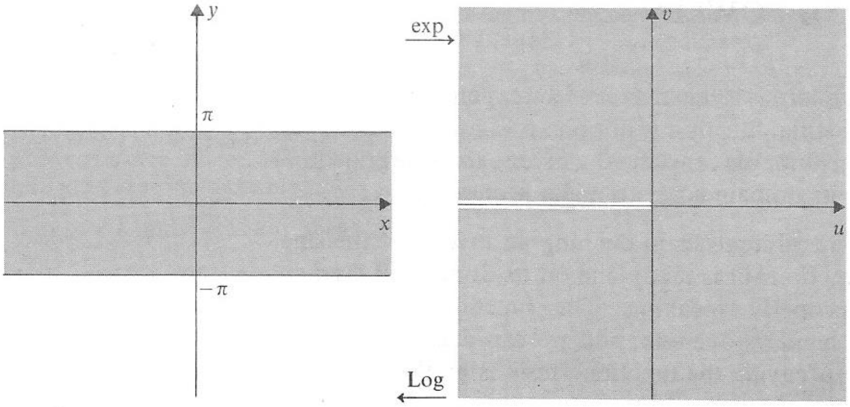


Fig. 18

Definition

> The exponential function with domain restricted to $\{z: -\pi < \operatorname{Im} z < \pi\}$ is one-one and its inverse function is called the **principal branch of the logarithm**, and is denoted by Log. The domain of Log is $\{z: z$ is not zero or a negative real number$\}$.

(The significance of the term "principal branch" will become apparent later.) Thus

$$\operatorname{Log}(e^z) = z \quad \text{when } -\pi < \operatorname{Im} z < \pi,$$

$$e^{\operatorname{Log} z} = z \quad \text{when } z \text{ is not zero or a negative real number.}$$

We shall next derive a formula for Log z (in fact you have almost derived this formula already, in Problem 1 of Section 3.5). Let us write $\operatorname{Log} z = u + iv$. Then

$$e^{\operatorname{Log} z} = e^u(\cos v + i \sin v) = z.$$

So $|e^{\operatorname{Log} z}| = e^u = |z|$, and $u = \log |z|$ (real logarithm; of course $|z| > 0$). Moreover, v is an argument of z, and since we require that $-\pi < v < \pi$, v is in fact the principal value of the argument: $v = \operatorname{Arg} z$. Thus

$$\operatorname{Log} z = \log |z| + i \operatorname{Arg} z.$$

(So if z is a positive real number, Log z is just the ordinary real logarithm of z.)

You may have noticed that exp is one-one on the set $\{z: -\pi < \operatorname{Im} z \leqslant \pi\}$, which is a little larger than the region we have been discussing up to now, and wondered why we did not define Log as the inverse of exp on this larger set. The reason is this. The image of $\{z: -\pi < \operatorname{Im} z \leqslant \pi\}$ under exp is the whole complex plane with the exception of the origin only. The formula we have just derived for Log would still hold, and give $\operatorname{Log} z_0 = \log |z_0| + i\pi$ when z_0 is a negative real number (not zero). It is pretty clear that Log, defined like this, would not be continuous across the negative real axis, since a complex number z near z_0 but with negative imaginary part will have argument close to $-\pi$, and so Log z will differ from Log z_0 by nearly $2i\pi$.

However, Log is continuous on $\{z: z \neq 0$ or a negative real number$\}$. This is clear from the formula: $z \longrightarrow \log |z|$ and $z \longrightarrow \operatorname{Arg} z$ are both continuous. The latter assertion is geometrically very plausible, but the full *analytic* details are a little tedious so we omit them. Here is a *geometric* proof of the fact that Arg is continuous on $\{z: -\pi < \operatorname{Arg} z < \pi\}$.

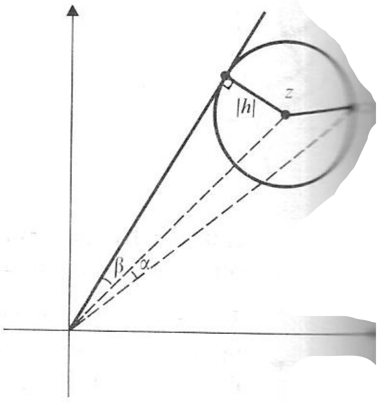


Fig. 19

From Fig. 19 we see that $\alpha \leqslant \beta$. Also, the circle centre z, where $z \in \{z: -\pi < \operatorname{Arg} z < \pi\}$ does not touch the non-positive real axis if $|h|$ is sufficiently small. So

$$|\operatorname{Arg}(z + h) - \operatorname{Arg} z| = \alpha \leqslant \beta = \arcsin\left(\frac{|h|}{|z|}\right)$$

provided $|h|$ is sufficiently small. But $\lim\limits_{h \to 0} \left(\frac{|h|}{|z|}\right) = 0$ and the result follows.

Since we now know that Log is continuous, we may use the theorem about the derivative of an inverse function (Theorem 7 of Section 3.1) to show that it is analytic.

Theorem 10

Log is analytic on its domain, and $\text{Log}' z = \frac{1}{z}$.

Proof

By Theorem 7 of Section 3.1, since Log is continuous everywhere in its domain, and $\exp'(z) = \exp z$ is never zero, we have

$$\text{Log}'(z) = \frac{1}{\exp'(\text{Log}\, z)} = \frac{1}{\exp(\text{Log}\, z)} = \frac{1}{z}. \quad \blacksquare$$

There is a slight problem associated with the function Log: it does not have the property $\text{Log}(z_1 z_2) = \text{Log}\, z_1 + \text{Log}\, z_2$. In the first place, it may happen that although z_1 and z_2 are both in the domain of Log, $z_1 z_2$ is not. For example, if $z_1 = z_2 = i$, then $z_1 z_2 = -1$. Secondly, we may fall foul of the restriction $-\pi < \text{Im}(\text{Log}\, z) < \pi$. For example, if $z_1 = z_2 = i - 1$, $\text{Arg}\, z_1 = \text{Arg}\, z_2 = 3\pi/4$, and $\text{Log}\, z_1 + \text{Log}\, z_2 = \log 2 + i3\pi/2$ which cannot be the value of the Log of any complex number. These problems are in a way technical problems: in other words, it is more or less true that $\text{Log}(z_1 z_2) = \text{Log}\, z_1 + \text{Log}\, z_2$; provided you keep a cool head, you will be able to deal on an *ad hoc* basis with any multiples of $2i\pi$ that crop up.

Theorem 11

If none of z_1, z_2 and $z_1 z_2$ is a nonpositive real number, then

$$\text{Log}(z_1 z_2) = \text{Log}\, z_1 + \text{Log}\, z_2 + 2ni\pi, \quad \text{where } n = -1, 0, \text{ or } 1.$$

Proof

$$\begin{aligned}\text{Log}(z_1 z_2) &= \log |z_1 z_2| + i\,\text{Arg}(z_1 z_2)\\ &= \log |z_1| + \log |z_2| + i(\text{Arg}\, z_1 + \text{Arg}\, z_2 + 2n\pi)\\ &= \text{Log}\, z_1 + \text{Log}\, z_2 + 2ni\pi.\end{aligned}$$

In the equation $\text{Arg}(z_1 z_2) = \text{Arg}\, z_1 + \text{Arg}\, z_2 + 2n\pi$, the integer n can take only the values -1, 0, or 1. ■

Let us summarize what we have done so far. We have pointed out that exp cannot have an inverse, since it is not one-one; we have constructed the inverse function for the function obtained by restricting the domain of exp to the region $\{z: -\pi < \text{Im}\, z < \pi\}$, which we called the principal branch of the logarithm, Log; we have shown that Log has almost all the properties one would expect of a logarithm function.

Our discussion has depended on choosing a particular region on which exp is one-one, and there are many such regions (see Problem 2 of Section 3.5). There are consequently many possible logarithm functions. We adopt the following terminology.

Definition

A region R such that exp is one-one on R, but on no region properly containing R, is called a **fundamental region for** exp. The inverse of the restriction of exp to a fundamental region R is called the **branch of the logarithm associated with** R, and is denoted by L_R.

So L_R satisfies

$$\exp(L_R(w)) = w;$$
$$L_R(w) \in R.$$

The domain of L_R is $\exp(R)$, the image of R under exp. It is always a region, though this is not immediately obvious. The name "branch of the logarithm", which may strike you as a bit odd, is standard terminology in complex analysis. It arose because one may describe the one-many mapping which is the inverse of exp as "the logarithm". A one-many mapping is of little use in analysis, though, so one has to divide it up into component functions, or "branches".

We have defined $L_R(z)$: we now show how to calculate it. As an example of a fundamental region for exp, consider the region $R = \{z : 0 < \operatorname{Im} z < 2\pi\}$. The image of this region under exp is the complex plane with the positive real axis and 0 removed.

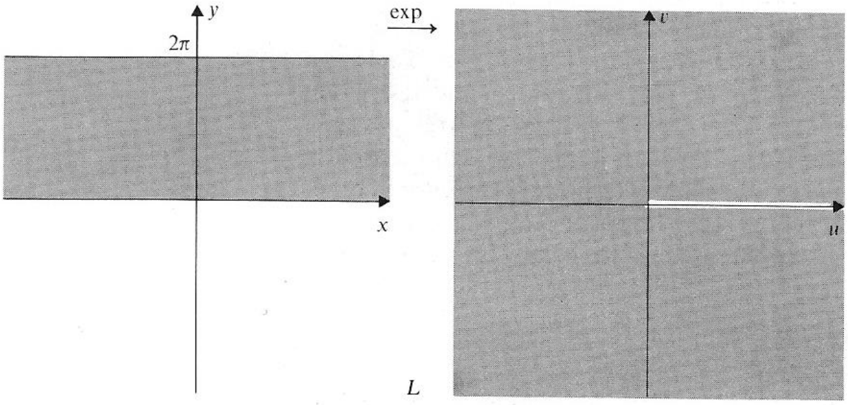


Fig. 20

The inverse of exp restricted to this fundamental region is a branch of the logarithm which we shall denote by L_R. It must satisfy

$$\exp(L_R(z)) = z, \quad z \neq 0 \text{ or a positive real number;}$$

and also

$$L_R(z) \in R \quad \text{so that} \quad 0 < \operatorname{Im} L_R(z) < 2\pi.$$

It follows from the first of these conditions that $L_R(z) = \log|z| + i\theta(z)$, where $\theta(z)$ is an argument of z; and the second condition tells us to choose that argument of z which lies in $(0, 2\pi)$.

The relation between this branch of the logarithm and the principal branch is as follows:

$$L_R(z) = \operatorname{Log} z \qquad \text{if} \quad 0 < \operatorname{Arg} z < \pi;$$
$$L_R(z) = \operatorname{Log} z + 2i\pi, \quad \text{if} \; -\pi < \operatorname{Arg} z < 0.$$

In other words, the two branches agree on the upper half-plane, but differ on the lower half-plane. The new branch is not defined for positive real numbers; but it is defined for negative real numbers, and if z is a negative real number, $L_R(z) = \log|z| + i\pi$.

Incidentally, we had logarithms in mind when we chose as principal value of the argument of a complex number that argument lying in $(-\pi, \pi]$. The intention was to choose the simplest formula for the principal branch of the logarithm. We are free to choose which branch of the logarithm to call the principal branch, of course; but a natural choice is one which contains the positive real axis in its domain, and agrees with the ordinary logarithm there.

There are many other branches of the logarithm. Some have the same domain as the principal branch: for example, that defined by the fundamental region $\{z : 3\pi < \operatorname{Im} z < 5\pi\}$. Some have quite different domains: for example, that branch associated with the fundamental region $\{z:\pi/6 < \operatorname{Im} z < 13\pi/6\}$. The calculation of $L_R(z)$ for a given region R is straightforward. We know that it must be $\log|z| + i\theta(z)$ where $\theta(z)$ is an argument of z. We simply choose that argument which ensures that the value of $L_R(z)$ lies in R.

Note that from here on we use L for some branch of the logarithm where R is not specifically defined.

Self-Assessment Questions

1. (a) Evaluate (i) $\operatorname{Log} i$; (ii) $\operatorname{Log}(1-i)$.

   (b) Evaluate (i) $L_R(i)$; (ii) $L_R(1-i)$, where $R = \{z:0 < \operatorname{Im} z < 2\pi\}$.

   (c) Evaluate $\operatorname{Log}(1+i)$ by using Theorem 11. (Note: $1 + i = i(1-i)$.)

2. Is it true that $\operatorname{Log}(1/z) = -\operatorname{Log} z$ for every z in the domain of Log?

Solutions

1. (a) (i) $\operatorname{Log} i = \log|i| + i\operatorname{Arg} i = \log 1 + i\dfrac{\pi}{2} = i\dfrac{\pi}{2}$.

   (ii) $\operatorname{Log}(1-i) = \log|1-i| + i\operatorname{Arg}(1-i) = \log\sqrt{2} + i\left(-\dfrac{\pi}{4}\right)$.

   (b) (i) $L_R(i) = \log|i| + i\theta(i)$ where $\theta(i)$ is the argument of i in R. So $L_R(i) = 0 + i\dfrac{\pi}{2} = i\dfrac{\pi}{2}$.

   (ii) $L_R(1-i) = \log|1-i| + i\theta(1-i)$ where $\theta(1-i)$ is the argument of $1-i$ in R. So $L_R(1-i) = \log\sqrt{2} + i\dfrac{7\pi}{4}$.

   (c)
$$\begin{aligned}\operatorname{Log}(1+i) &= \operatorname{Log}(i(1-i))\\ &= \operatorname{Log} i + \operatorname{Log}(1-i) + 2ni\pi, \quad \text{by Theorem 11,}\\ &= \log\sqrt{2} + i\frac{\pi}{4} + 2ni\pi\\ &= \log\sqrt{2} + i\frac{\pi}{4}, \quad \text{since } \frac{\pi}{4}\in(-\pi,\pi).\end{aligned}$$

2. It is true that $\operatorname{Log}(1/z) = -\operatorname{Log} z$ for every z in the domain of Log. Notice first that if z is not a nonpositive real number, neither is $1/z$, so the equation makes sense. Now

$$\operatorname{Log}\frac{1}{z} = \log\frac{1}{|z|} + i\operatorname{Arg}\frac{1}{z};$$

$\log\dfrac{1}{|z|} = -\log|z|$ and $\operatorname{Arg}\dfrac{1}{z} = -\operatorname{Arg} z$ since z is not a negative real number. Thus

$$\begin{aligned}\operatorname{Log}\frac{1}{z} &= -\log|z| - i\operatorname{Arg} z\\ &= -\operatorname{Log} z.\end{aligned}$$

Powers

Definition

Let α be a complex number and L_R a branch of the logarithm associated with a fundamental region R. The function $z \longrightarrow \exp(\alpha L_R(z))$ for z in the domain of L_R, is called a **power function.** We also use the term a **branch of the αth power** if we wish to emphasize that a choice of branch of the logarithm is involved. The function $z \longrightarrow \exp(\alpha \operatorname{Log} z)$, where we use the principal branch of the logarithm, is called the **principal branch of the αth power.**

This definition should bring to mind the definition of x^a when x is a positive (real) number and a an irrational (real) number: we define x^a to be $e^{a \log x}$.

If n is an integer, and L any branch of the logarithm, then $\exp(n(L(z)) = [\exp(L(z))]^n = z^n$; so a branch of the nth power is to all intents and purposes just the same as the function $z \longrightarrow z^n$. This justifies our choice of the name "power function", and explains the purpose of the definition: to generalize the notion of nth power.

So if α is an integer nothing new of interest occurs. The next case, in order of increasing difficulty, is α a rational (real) number. We shall try to show what sort of thing happens by taking a particular example, $\alpha = \frac{1}{2}$.

The first thing to notice is that if L is a branch of the logarithm and z is in the domain of L, then $[\exp(\frac{1}{2}L(z))]^2 = \exp(L(z)) = z$. Thus $e^{L(z)/2}$ is a square root of z. In fact $\frac{1}{2}L(z) = \log\sqrt{|z|} + \frac{1}{2}i\theta(z)$, where $\theta(z)$ is an argument of z, and so

$$\begin{aligned}\exp(\tfrac{1}{2}L(z)) &= \exp(\log\sqrt{|z|})[\cos\tfrac{1}{2}\theta(z) + i\sin\tfrac{1}{2}\theta(z)]\\ &= \sqrt{|z|}[\cos\tfrac{1}{2}\theta(z) + i\sin\tfrac{1}{2}\theta(z)],\end{aligned}$$

which is the formula for a square root given in *Unit 1*, Section 1.3. We shall call $z \longrightarrow \exp(\frac{1}{2}L(z)) = e^{L(z)/2}$ a **branch of the square root**, rather than using the somewhat grotesque phrase "a branch of the $\frac{1}{2}$th power".

The domain of the principal branch of the square root contains the positive real axis, and if x is a positive real number $e^{(\log x)/2} = \sqrt{x}$ is the usual positive real square root. The principal branch of the square root extends the square root function of real analysis in a natural way into the complex plane. But it does not include the negative real axis in its domain; in fact, it is impossible to find a function defined on the whole complex plane, which is continuous everywhere, and whose value at each point z is a square root of z. If we wish to find a branch of the square root which has the negative real axis in its domain, we must choose an appropriate branch of the logarithm. One possibility is the branch corresponding to the fundamental region $S = \{z : 0 < \operatorname{Im} z < 2\pi\}$ for exp. If L_S denotes this branch, and z is a negative real number, then $L_S(z) = \log|z| + i\pi$. Thus $\exp(\frac{1}{2}L_S(z)) = \sqrt{|z|} \cdot e^{i\pi/2} = i\sqrt{|z|}$.

There are infinitely many branches of the square root, but it is important that you realize that there is no contradiction between this and the fact that a non-zero complex number has just two square roots. A branch of the square root is a function, after all; and the proliferation of these functions is mainly due to the different possibilities for their domains. We have shown above that the value of any branch of the square root at a point z (where $z \neq 0$) is one or other of the square roots of z.

More generally, if α is a rational real number, say $\alpha = p/q$, where p and q are integers, $q \neq 0$, and $z \neq 0$, the possible values of the branches of the αth power at z are just the q qth roots of z^p. Suppose that L is a branch of the logarithm whose domain contains z. Then

$$\begin{aligned}[\exp(\alpha L(z))]^q &= \exp(q\alpha L(z))\\ &= \exp(pL(z))\\ &= [\exp(L(z))]^p\\ &= z^p.\end{aligned}$$

So $\exp(\alpha(L(z))$ is indeed a qth root of z^p. The possible values of the branches of the αth power at z are the numbers $\{\exp(\alpha[L(z)+2ki\pi]) : k=0, \pm 1, \pm 2, \ldots\}$. Now

$$\exp(\alpha[L(z) + 2ki\pi]) = \exp(\alpha[L(z) + 2li\pi])$$

if and only if

$$\exp(\alpha 2(k - l)i\pi) = 1$$

that is, if and only if $\alpha 2(k - l)i\pi$ is an integer multiple of $2i\pi$. This occurs only when k and l differ by a multiple of q. So the possible distinct values of the branches of the αth power at z are the numbers $\{\exp(\alpha[L(z) + 2ki\pi]) : k = 0, 1, 2, \ldots, q - 1\}$, and these are just the q distinct qth roots of z^p.

If α is not rational on the other hand, the numbers $\{\exp(\alpha[L(z) + 2ki\pi]) : k = 0, \pm 1, \pm 2, \ldots\}$ are all different, since in this case $\alpha 2(k - l)i\pi$ can only be an integer multiple of $2i\pi$ if $k = l$. We conclude that there are infinitely many distinct values of the branches of the αth power at z when α is not a rational number.

We have been careful not to use the symbol z^α except when α is an integer. There is not in general a unique complex number which can be thought of as z^α. We shall, however, sometimes use z^α to denote all numbers of the form $e^{\alpha L(z)}$, where L is a branch of the logarithm. So if n is a positive integer, $z^{1/n}$ means all n nth roots of z; but if α is not a rational real number, there are infinitely many numbers z^α. The number $e^{\alpha \operatorname{Log} z}$ is called the **principal value of** z^α.

Examples

1. Show that the numbers $z^{1/n}$ are obtained by multiplying a particular nth root of z by the n nth roots of 1.

   Let L be a branch of the logarithm and let z be in the domain of L. Then the possible values of $z^{1/n}$ are $\{e^{[L(z)+2mi\pi]/n} : m \text{ an integer}\}$. We need distinguish only those powers which do not differ by an integer multiple of $2i\pi$, so a suitable set of distinct values of $z^{1/n}$ is

   $$\begin{aligned}&\{e^{[L(z)+2mi\pi]/n} : m = 0, 1, 2, \ldots, n-1\}\\ &= \{e^{L(z)/n}, e^{L(z)/n}\cdot e^{(2i\pi)/n}, e^{L(z)/n}\cdot e^{(4i\pi)/n}, \ldots, e^{L(z)/n}\cdot e^{[2(n-1)\overset{i}{\wedge}\pi]/n}\}.\end{aligned}$$

   But $e^{L(z)/n}$ is a particular nth root of z, and $\{1, e^{(2i\pi)/n}, e^{(4i\pi)/n}, \ldots, e^{[2(n-1)\overset{i}{\wedge}\pi]/n}\}$ are the n nth roots of 1.

2. Find the possible values of $(1 + i)^{2/3}$, and its principal value.

   We have

   $$\operatorname{Log}(1 + i) = \log\sqrt{2} + i\frac{\pi}{4},$$

   so

   $$\exp\left(\frac{2}{3}\operatorname{Log}(1 + i)\right) = \sqrt[3]{2}e^{i\pi/6} = \sqrt[3]{2}\left(\frac{\sqrt{3}}{2} + i\frac{1}{2}\right).$$

   This is the principal value of $(1 + i)^{2/3}$. Its other distinct values are

$$\exp(\tfrac{2}{3}[\mathrm{Log}(1+i)+2i\pi]) \quad \text{and} \quad \exp(\tfrac{2}{3}[\mathrm{Log}(1+i)+4i\pi]),$$

which give

$$\sqrt[3]{2}e^{9i\pi/6} = \sqrt[3]{2}e^{3i\pi/2} = -\sqrt[3]{2}\cdot i$$

and

$$\sqrt[3]{2}e^{17i\pi/6} = \sqrt[3]{2}e^{5i\pi/6} = \sqrt[3]{2}\left(-\frac{\sqrt{3}}{2}+i\frac{1}{2}\right).$$

3. What is the principal value of i^i?

$\mathrm{Log}\, i = i\dfrac{\pi}{2}$. Thus $e^{i\,\mathrm{Log}\, i} = e^{-\pi/2}$. The principal value of i^i is $e^{-\pi/2}$.

4. Find all possible values of $(1+i)^{(1+i)}$.

The possible values of $(1+i)^{(1+i)}$ are

$$\begin{aligned}
&\exp\left((1+i)\left(\log\sqrt{2}+i\frac{\pi}{4}+2ni\pi\right)\right), \quad n = 0, \pm 1, \pm 2, \ldots\\
&= \exp\left(\left[\log\sqrt{2}-\left(2n+\frac{1}{4}\right)\pi\right]+i\left[\log\sqrt{2}+\left(2n+\frac{1}{4}\right)\pi\right]\right)\\
&= \exp\left(\left[\log\sqrt{2}-\left(2n+\frac{1}{4}\right)\pi\right]+i\left[\log\sqrt{2}+\frac{\pi}{4}\right]\right)\\
&= \sqrt{2}\exp\left(-\left(2n+\frac{1}{4}\right)\pi\right)\left[\cos\left(\log\sqrt{2}+\frac{\pi}{4}\right)+i\sin\left(\log\sqrt{2}+\frac{\pi}{4}\right)\right],\\
&\qquad n = 0, \pm 1, \pm 2, \ldots .
\end{aligned}$$

Self-Assessment Questions

3. Show that the power function $z \longrightarrow e^{\alpha L(z)}$ (where L is a branch of the logarithm) is analytic on its domain, and find its derivative.
4. Find all possible values of i^i.

Solutions

3. The function $z \longrightarrow e^{\alpha L(z)}$ is analytic on its domain since it is the composition of analytic functions. Also

$$\frac{d}{dz}e^{\alpha L(z)} = \alpha\cdot\frac{1}{z}\cdot e^{\alpha L(z)} = \alpha e^{(\alpha-1)L(z)}.$$

4. $\exp\left(-\dfrac{\pi}{2}+2n\pi\right), \quad n = 0, \pm 1, \pm 2, \ldots$

Summary

We found difficulty in defining a complex logarithm function as the inverse of the complex exponential function because the latter function is not one-one. We got over this difficulty by restricting the domain of exp (much as we define arcsin in real analysis by restricting the domain of sin). There are consequently many possible logarithm functions; they are called branches of the logarithm.

There are many branches of the αth power function, too; if $\alpha = 1/n$, these just give the n nth roots.

3.7 PROBLEMS

1. What is the domain of the function $z \longrightarrow \mathrm{Log}(1+z^2)$? Where is it analytic? What is its derivative?
2. Show that if L is a branch of the logarithm and z lies in both the domain of L and the domain of Log, then $L(z) = \mathrm{Log}\, z + 2n i\pi$, where n is an integer.
3. It is claimed in the text that $S = \{z : \pi/6 < \mathrm{Im}\, z < 13\pi/6\}$ is a fundamental region for exp. Show that this is correct. What is the domain of the corresponding branch of the logarithm? Find
   (i) $L_S(1)$; (ii) $L_S(-1)$; (iii) $L_S(i)$; (iv) $L_S(-i)$.
4. Find all the possible values of
   (i) $(-i)^i$; (ii) $(1-i)^{(1+i)}$.
5. Show that $\{z : \mathrm{Re}\, z > 0\}$ is a fundamental region for the function $z \longrightarrow z^2$. (Part of the problem is to guess what the phrase "a fundamental region for the function $z \longrightarrow z^2$" means.) Show that the inverse function of $z \longrightarrow z^2$ on $\{z : \mathrm{Re}\, z > 0\}$ is the principal branch of the square root.

Solutions

1. The domain of $z \longrightarrow \mathrm{Log}(1+z^2)$ is $\{z : 1 + z^2 \neq 0$ or a negative real number$\}$. To find what this is explicitly, we find those complex numbers z for which $1 + z^2$ is 0 or a negative real number; the set we want consists of all other complex numbers. If $z = x + iy$, then $1 + z^2 = 1 + x^2 - y^2 + 2ixy$. This can be real only if $x = 0$ or $y = 0$. If $y = 0$, the real part of $1 + z^2$ is positive. So we must have $x = 0$, and $y^2 \geqslant 1$, in order that $1 + z^2$ be a nonpositive real number. So the domain of $z \longrightarrow \mathrm{Log}(1+z^2)$ is the complex plane with the exception of those points on the imaginary axis for which $|z| \geqslant 1$ (Fig. 21).

   The function is analytic on its domain, and its derivative is $z \longrightarrow \dfrac{2z}{1+z^2}$ by the Chain Rule.

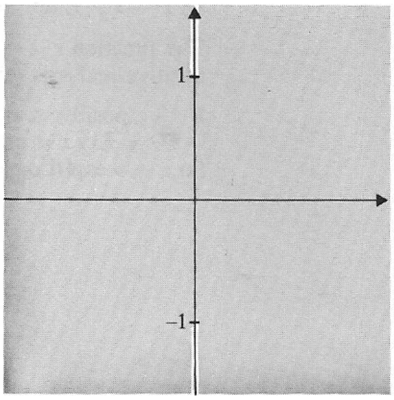


Fig. 21

2. $e^{L(z)} = e^{\mathrm{Log}\, z} = z$, so $e^{L(z) - \mathrm{Log}\, z} = 1$. But this implies that $L(z) - \mathrm{Log}\, z = 2n i\pi$ for some integer n.
3. The set $S = \{z : \pi/6 < \mathrm{Im}\, z < 13\pi/6\}$ is certainly a region, and if it contains z then it contains no complex number which differs from z by a multiple of $2i\pi$. Thus exp is one-one on the region. Clearly any larger region must contain points differing from points of the given region by multiples of $2i\pi$. So the region is indeed a fundamental region for exp.

   The image of this region under exp is the complex plane with the ray $\{z : \mathrm{Arg}\, z = \pi/6\}$, and the origin, excluded. This is the domain of L_S, the corresponding branch of the logarithm.

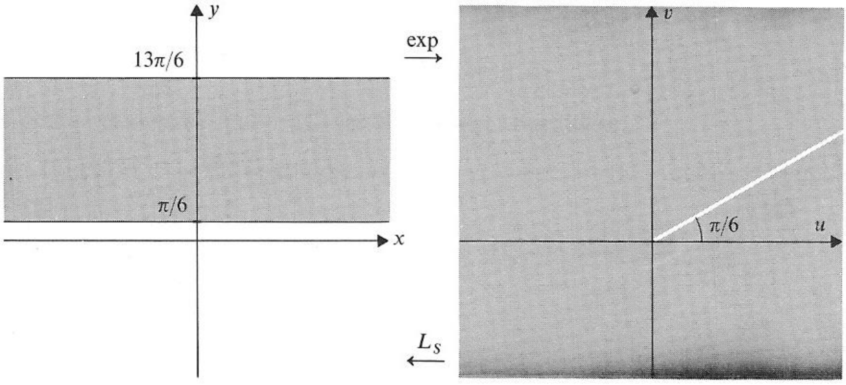


Fig. 22

$L_S(z) = \log|z| + i\theta(z)$, where $\theta(z)$ is the argument of z such that $\pi/6 < \theta(z) < 13\pi/6$.

(i) $L_S(1) = \log 1 + 2ni\pi$, where n is such that $\pi/6 < 2ni\pi < 13\pi/6$. Thus $L_S(1) = 2i\pi$.

(ii) $L_S(-1) = i(\pi + 2n\pi)$, where n is such that $\pi/6 < \pi + 2n\pi < 13\pi/6$. Thus $L_S(-1) = i\pi$.

(iii) $L_S(i) = \dfrac{i\pi}{2}$.

(iv) $L_S(-i) = \dfrac{3i\pi}{2}$.

4. (i) $\exp(\pi/2 + 2n\pi)$; $n = 0, \pm 1, \pm 2, \ldots$.

(ii) $\sqrt{2}\exp\big((2n + \tfrac{1}{4})\pi\big)[\cos(\log\sqrt{2} - \pi/4) + i\sin(\log\sqrt{2} - \pi/4)]$,

$n = 0, \pm 1, \pm 2, \ldots$.

5. A fundamental region for the function $z \longrightarrow z^2$ is a region R such that $z \longrightarrow z^2$ is one-one on R, but not one-one on any region properly containing R. (Simply substitute $z \longrightarrow z^2$ for exp in the definition of a fundamental region for exp.)

The set $\{z : \operatorname{Re} z > 0\}$ is certainly a region on which $z \longrightarrow z^2$ is one-one (since $z_1^2 = z_2^2$ if and only if $z_2 = \pm z_1$). If S is any larger region then it must contain a number z_1 with $\operatorname{Re} z_1 < 0$ (since there is no region contained in $\operatorname{Re} z \geqslant 0$ apart from $\operatorname{Re} z > 0$ and its subregions). But then if $z_2 = -z_1$, $\operatorname{Re} z_2 > 0$, and S contains both z_1 and z_2. So $z \longrightarrow z^2$ is not one-one on S.

The function $z \longrightarrow z^2$ maps $\{z : \operatorname{Re} z > 0\}$ onto the complex plane with the non-positive real axis removed.

If z is a positive real number, then $[\exp(\frac{1}{2}\operatorname{Log} z)]^2 = \exp(\operatorname{Log} z) = z$. If $\operatorname{Re} z > 0$, then $-\pi/2 < \operatorname{Arg} z < \pi/2$, so $\operatorname{Log}(z^2) = 2\operatorname{Log} z$. Thus $\exp([\operatorname{Log}(z^2)]/2) = \exp(\operatorname{Log} z) = z$. So $z \longrightarrow \exp\big((\operatorname{Log} z)/2\big)$ is the inverse of the function $z \longrightarrow z^2$ $(\operatorname{Re} z > 0)$.

3.8 THE CAUCHY–RIEMANN EQUATIONS

Partial Differentiation: definitions and results

In this section we shall be discussing functions defined on $\mathbf{R}^2$ and we shall need a number of results on partial differentiation. If you are not familiar with partial differentiation you can gain the appropriate knowledge (which is specified below) by consulting certain appendices to M231 units (see below) or by consulting L. Leithold, *The Calculus with analytic geometry* (Harper and Row, 1972).

So far in the course we have emphasized the similarities between complex analysis and real analysis, which rest on the similarities between the complex and real number systems. In the rest of this unit we take a different point of view. The complex numbers and the set of pairs of real numbers are identical as sets: they are identical, in other words, if we ignore some of the structure of the complex numbers. Looked at from this point of view, complex analysis is just a special case of the analysis of functions from $\mathbf{R}^2$ to $\mathbf{R}^2$. In practice we cannot get very far if we adopt this attitude consistently; and in many ways it is aesthetically less pleasing than the intrinsic approach we have used up to now. Nevertheless there are useful insights, and practical advantages, to be gained from changing our point of view from time to time.

Since we are going to be discussing functions defined on $\mathbf{R}^2$, we shall need to use partial differentiation. We have made a collection of the main definitions and results, which you may use for reference. (The proofs of these results are to be found in the appendices to *Unit M231 3, Limits and Continuity*, *Unit M231 4, Continuous Functions*, and *Unit M231 5, Differentiation.*)

Definition

The function $F:\mathbf{R}^2 \longrightarrow \mathbf{R}$ is **continuous at** (α, β) if for every $\varepsilon > 0$ there is some $\delta > 0$ such that

$$\text{if } 0 < \sqrt{(x-\alpha)^2 + (y-\beta)^2} < \delta \text{ then } |F(x, y) - F(\alpha, \beta)| < \varepsilon.$$

Remember that $\sqrt{(x-\alpha)^2 + (y-\beta)^2}$ is the *distance* from (x, y) to (α, β); if we identify (x, y) and (α, β) with complex numbers $z = x + iy$ and $a = \alpha + i\beta$, then $\sqrt{(x-\alpha)^2 + (y-\beta)^2} = |z - a|$. This definition of continuity is essentially the same as the definition of continuity we use for complex functions.

Definition

If $\lim\limits_{h\to 0} \dfrac{F(\alpha + h, \beta) - F(\alpha, \beta)}{h}$ exists, it is called the **partial derivative of** F **at** (α, β) **for the first coordinate**; it is denoted by $\dfrac{\partial F}{\partial x}(\alpha, \beta)$. Similarly if $\lim\limits_{k\to 0} \dfrac{F(\alpha, \beta + k) - F(\alpha, \beta)}{k}$ exists, we denote it by $\dfrac{\partial F}{\partial y}(\alpha, \beta)$ and call it the **partial derivative of** F **at** (α, β) **for the second coordinate**.

The function F is **differentiable at** (α, β) if there are numbers p, q such that if

$$\Phi(h, k) = F(\alpha + h, \beta + k) - F(\alpha, \beta) - p \cdot h - q \cdot k,$$

then

$$\lim_{(h,k)\to(0,0)} \frac{\Phi(h, k)}{\sqrt{h^2 + k^2}} = 0.$$

Theorem 12

If F is differentiable at (α, β) then its partial derivatives exist at (α, β), and in fact $p = \dfrac{\partial F}{\partial x}(\alpha, \beta)$ and $q = \dfrac{\partial F}{\partial y}(\alpha, \beta)$.

We are forced into this slightly round about way of doing things by the unfortunate fact that a function may have partial derivatives at a point without being differentiable there. However, if the partial derivatives exist in a neighbourhood of (α, β), and are continuous at (α, β), then F is differentiable there.

Theorem 13

If there is a neighbourhood of (α, β) such that $\dfrac{\partial F}{\partial x}(x, y)$ and $\dfrac{\partial F}{\partial y}(x, y)$ both exist for all (x, y) in this neighbourhood, and if $\dfrac{\partial F}{\partial x}$ and $\dfrac{\partial F}{\partial y}$ are both continuous at (α, β), then F is differentiable at (α, β).

The Cauchy–Riemann Equations

When we proved that $z \longrightarrow \bar{z}$ is not differentiable we did it by computing $\dfrac{\overline{(z + h)} - \bar{z}}{h}$ when h is real, and when h is imaginary. We got values which showed that $\dfrac{\overline{(z + h)} - \bar{z}}{h}$ cannot have a limit as h approaches 0.

Conversely, if a function f is known to be differentiable at a point z_0, then by examining $\dfrac{f(z_0 + h) - f(z_0)}{h}$ for real h and for imaginary h we obtain some interesting information about the function f. This information can be expressed as relations between the partial derivatives of the real and imaginary parts of f. These relations, which are given in Theorem 14, are called the **Cauchy–Riemann equations**.

Theorem 14

If the function f is differentiable at the point $z_0 = x_0 + iy_0$, and $f(x + iy) = U(x, y) + iV(x, y)$, where U and V (the real and imaginary parts of f) are considered as real-valued functions of two real variables, then U and V have partial derivatives at (x_0, y_0) which satisfy the Cauchy-Riemann equations

$$\frac{\partial U}{\partial x}(x_0, y_0) = \frac{\partial V}{\partial y}(x_0, y_0); \quad \frac{\partial V}{\partial x}(x_0, y_0) = -\frac{\partial U}{\partial y}(x_0, y_0).$$

Also

$$f'(z_0) = \frac{\partial U}{\partial x}(x_0, y_0) + i\frac{\partial V}{\partial x}(x_0, y_0) = \frac{\partial V}{\partial y}(x_0, y_0) - i\frac{\partial U}{\partial y}(x_0, y_0).$$

Proof

Since f is differentiable at z_0, given any $\varepsilon > 0$ there is some $\delta > 0$ such that

$$\left|\frac{f(z_0 + h) - f(z_0)}{h} - f'(z_0)\right| < \varepsilon \quad \text{whenever } 0 < |h| < \delta.$$

So if α is real and $0 < |\alpha| < \delta$,

$$\left|\frac{f(z_0 + \alpha) - f(z_0)}{\alpha} - f'(z_0)\right| < \varepsilon.$$

But

$$\frac{f(z_0+\alpha)-f(z_0)}{\alpha}=\frac{U(x_0+\alpha, y_0)-U(x_0, y_0)}{\alpha}+i\,\frac{V(x_0+\alpha, y_0)-V(x_0, y_0)}{\alpha}$$

and since $|\mathrm{Re}\,\zeta| \leqslant |\zeta|$ and $|\mathrm{Im}\,\zeta| \leqslant |\zeta|$ for any complex number ζ, we conclude that

$$\left|\frac{U(x_0+\alpha, y_0)-U(x_0, y_0)}{\alpha}-\mathrm{Re}\,f'(z_0)\right|<\varepsilon$$

and

$$\left|\frac{V(x_0+\alpha, y_0)-V(x_0, y_0)}{\alpha}-\mathrm{Im}\,f'(z_0)\right|<\varepsilon,$$

whenever $0<|\alpha|<\delta$. Thus $\dfrac{\partial U}{\partial x}(x_0, y_0)=\mathrm{Re}\,f'(z_0)$ and $\dfrac{\partial V}{\partial x}(x_0, y_0)=\mathrm{Im}\,f'(z_0)$.

On the other hand, if $i\beta$ is imaginary and $0<|i\beta|<\delta$ (or equivalently $0<|\beta|<\delta$), then

$$\left|\frac{f(z_0+i\beta)-f(z_0)}{i\beta}-f'(z_0)\right|<\varepsilon.$$

But

$$\frac{f(z_0+i\beta)-f(z_0)}{i\beta}=\frac{V(x_0, y_0+\beta)-V(x_0, y_0)}{\beta}-i\,\frac{U(x_0, y_0+\beta)-U(x_0, y_0)}{\beta}.$$

We conclude that

$$\left|\frac{V(x_0, y_0+\beta)-V(x_0, y_0)}{\beta}-\mathrm{Re}\,f'(z_0)\right|<\varepsilon$$

and

$$\left|-\frac{U(x_0, y_0+\beta)-U(x_0, y_0)}{\beta}-\mathrm{Im}\,f'(z_0)\right|<\varepsilon,$$

whenever $0<|\beta|<\delta$. Thus $\dfrac{\partial V}{\partial y}(x_0, y_0)=\mathrm{Re}\,f'(z_0)$ and $\dfrac{\partial U}{\partial y}(x_0, y_0)=-\mathrm{Im}\,f'(z_0)$.

Comparing these two conclusions, we deduce that

$$\frac{\partial U}{\partial x}(x_0, y_0)=\frac{\partial V}{\partial y}(x_0, y_0) \quad\text{and}\quad \frac{\partial V}{\partial x}(x_0, y_0)=-\frac{\partial U}{\partial y}(x_0, y_0). \quad \blacksquare$$

The next question is: if the real and imaginary parts of a function satisfy the Cauchy-Riemann equations, is the function differentiable? It turns out that an extra hypothesis is needed before we can conclude that the function is differentiable. The most useful (though not the strongest) theorem is the following one.

Theorem 15

Suppose that $f=U+iV$, where the functions $\mathrm{Re}\,f=U$ and $\mathrm{Im}\,f=V$ have partial derivatives which are continuous at (x_0, y_0) and satisfy the Cauchy-Riemann equations there. Then f is differentiable at $z_0=x_0+iy_0$, and its derivative is

$$f'(z_0)=\frac{\partial U}{\partial x}(x_0, y_0)+i\,\frac{\partial V}{\partial x}(x_0, y_0)=\frac{\partial V}{\partial y}(x_0, y_0)-i\,\frac{\partial U}{\partial y}(x_0, y_0).$$

The proof of this result is optional, but you should remember the result and be able to use it.

Proof

Since U and V have partial derivatives which are continuous at (x_0, y_0), they are differentiable there. Thus

$$U(x_0 + \alpha, y_0 + \beta) = U(x_0, y_0) + \alpha \cdot \frac{\partial U}{\partial x}(x_0, y_0) + \beta \cdot \frac{\partial U}{\partial y}(x_0, y_0) + \rho(\alpha, \beta)$$

$$V(x_0 + \alpha, y_0 + \beta) = V(x_0, y_0) + \alpha \cdot \frac{\partial V}{\partial x}(x_0, y_0) + \beta \cdot \frac{\partial V}{\partial y}(x_0, y_0) + \sigma(\alpha, \beta)$$

where

$$\lim_{(\alpha,\beta)\to(0,0)} \frac{\rho(\alpha, \beta)}{\sqrt{\alpha^2 + \beta^2}} = \lim_{(\alpha,\beta)\to(0,0)} \frac{\sigma(\alpha, \beta)}{\sqrt{\alpha^2 + \beta^2}} = 0.$$

Thus, if $h = \alpha + i\beta$, we have

$$\begin{aligned} f(z_0 + h) &= U(x_0 + \alpha, y_0 + \beta) + iV(x_0 + \alpha, y_0 + \beta) \\ &= U(x_0, y_0) + iV(x_0, y_0) + \alpha \cdot \left(\frac{\partial U}{\partial x}(x_0, y_0) + i\frac{\partial V}{\partial x}(x_0, y_0)\right) \\ &\quad + \beta \cdot \left(\frac{\partial U}{\partial y}(x_0, y_0) + i\frac{\partial V}{\partial y}(x_0, y_0)\right) + \rho(\alpha, \beta) + i\sigma(\alpha, \beta). \end{aligned}$$

Now $U(x_0, y_0) + iV(x_0, y_0) = f(z_0)$, of course; and since the partial derivatives of U and V satisfy the Cauchy-Riemann equations,

$$\begin{aligned} \frac{\partial U}{\partial y}(x_0, y_0) + i\frac{\partial V}{\partial y}(x_0, y_0) &= -\frac{\partial V}{\partial x}(x_0, y_0) + i\frac{\partial U}{\partial x}(x_0, y_0) \\ &= i\left(\frac{\partial U}{\partial x}(x_0, y_0) + i\frac{\partial V}{\partial x}(x_0, y_0)\right). \end{aligned}$$

So if we set $\zeta = \frac{\partial U}{\partial x}(x_0, y_0) + i\frac{\partial V}{\partial x}(x_0, y_0)$, we have

$$\begin{aligned} f(z_0 + h) &= f(z_0) + \alpha \cdot \zeta + \beta \cdot i\zeta + \rho(\alpha, \beta) + i\sigma(\alpha, \beta) \\ &= f(z_0) + h \cdot \zeta + \rho(\alpha, \beta) + i\sigma(\alpha, \beta) \\ &= f(z_0) + h \cdot \zeta + \phi(h), \end{aligned}$$

where we have written $\phi(h)$ for $\rho(\alpha, \beta) + i\sigma(\alpha, \beta)$.

By Theorem 8 (of Section 3.2), the theorem will be proved if we show that $\lim_{h\to 0} \frac{\phi(h)}{h} = 0$.

Now

$$\begin{aligned} \left|\frac{\phi(h)}{h}\right| &= \left|\frac{\rho(\alpha, \beta) + i\sigma(\alpha, \beta)}{h}\right| \leqslant \left|\frac{\rho(\alpha, \beta)}{h}\right| + \left|\frac{\sigma(\alpha, \beta)}{h}\right| \\ &= \frac{|\rho(\alpha, \beta)|}{\sqrt{\alpha^2 + \beta^2}} + \frac{|\sigma(\alpha, \beta)|}{\sqrt{\alpha^2 + \beta^2}}. \end{aligned}$$

But

$$\lim_{(\alpha,\beta)\to(0,0)} \frac{\rho(\alpha, \beta)}{\sqrt{\alpha^2 + \beta^2}} = \lim_{(\alpha,\beta)\to(0,0)} \frac{\sigma(\alpha, \beta)}{\sqrt{\alpha^2 + \beta^2}} = 0,$$

so $\lim_{h\to 0} \frac{\phi(h)}{h} = 0$. Thus f is differentiable at z_0. ■

One of the most obvious and useful consequences of these two theorems is a new technique for telling whether or not a function is differentiable, and finding its derivative if it is. We illustrate this technique with some examples which we have previously tackled by other methods.

Examples

1. $f(z) = z^2$.

   If $z = x + iy$, then $f(z) = x^2 - y^2 + 2ixy$, so the real and imaginary parts are

   $$U(x, y) = x^2 - y^2 \quad \text{and} \quad V(x, y) = 2xy.$$

   The partial derivatives are

   $$\frac{\partial U}{\partial x}(x, y) = 2x, \qquad \frac{\partial U}{\partial y}(x, y) = -2y,$$

   $$\frac{\partial V}{\partial x}(x, y) = 2y, \qquad \frac{\partial V}{\partial y}(x, y) = 2x.$$

   These functions are continuous everywhere, and satisfy the Cauchy–Riemann equations. Thus, by Theorem 15, f is differentiable and its derivative is $2x + i \cdot 2y = 2z$.

This example illustrates the method, but there is no particular advantage in using it there. If we look at the exponential function instead, we find a considerable saving in using the Cauchy–Riemann equations to find the derivative.

2. $f(z) = e^z$.

   Since $e^z = e^x(\cos y + i \sin y)$ by definition, the real and imaginary parts of e^z are

   $$U(x, y) = e^x \cos y \quad \text{and} \quad V(x, y) = e^x \sin y.$$

   The partial derivatives are

   $$\frac{\partial U}{\partial x}(x, y) = e^x \cos y, \qquad \frac{\partial U}{\partial y} = -e^x \sin y,$$

   $$\frac{\partial V}{\partial x}(x, y) = e^x \sin y \qquad \frac{\partial V}{\partial y} = e^x \cos y.$$

   These functions are continuous everywhere and satisfy the Cauchy–Riemann equations. Thus, by Theorem 15, f is differentiable and its derivative is $e^x \cos y + i \cdot e^x \sin y = e^z$.

Finally, if the Cauchy–Riemann equations are not satisfied, or if the partial derivatives fail to exist, then the function cannot be differentiable. As an example we use the modulus function.

3. $f(z) = |z|$.

   Since $f(z)$ is real for all z, $V = 0$; and $U(x, y) = \sqrt{x^2 + y^2}$. Now the partial derivatives of U do not exist at the origin, so f cannot be differentiable there.

   If $(x, y) \neq (0, 0)$, then

   $$\frac{\partial U}{\partial x}(x, y) = \frac{x}{\sqrt{x^2 + y^2}}, \qquad \frac{\partial U}{\partial y}(x, y) = \frac{y}{\sqrt{x^2 + y^2}},$$

   $$\frac{\partial V}{\partial x}(x, y) = 0, \qquad \frac{\partial V}{\partial y}(x, y) = 0;$$

   the Cauchy–Riemann equations are not satisfied, so f is differentiable nowhere.

The Cauchy–Riemann equations have important theoretical consequences too.

Theorem 16

If f is analytic on a region R and $f'(z) = 0$ for all z in R, then f is constant.

Proof

We first show that f must be locally constant on R: in other words, given any point z_0 in R, there is a disc centre z_0 contained in R on which f is constant—and then appeal to the result of Problem 4 of Section 2.8 of *Unit 2*, which says that a function locally constant on a region must be constant on the whole region.

Let $z_0 \in R$ and let D be an open disc centre z_0 contained in R (there must be such a disc since R is open). Now

$$f'(x+iy) = \frac{\partial U}{\partial x}(x,y) + i\frac{\partial V}{\partial x}(x,y) = \frac{\partial V}{\partial y}(x,y) - i\frac{\partial U}{\partial y}(x,y) = 0$$

for all $x + iy$ in R, and so

$$\frac{\partial U}{\partial x}(x,y) = \frac{\partial U}{\partial y}(x,y) = \frac{\partial V}{\partial x}(x,y) = \frac{\partial V}{\partial y}(x,y) = 0.$$

Since $\dfrac{\partial U}{\partial x}(x,y_0) = \dfrac{\partial V}{\partial x}(x,y_0) = 0$, the functions $x \longrightarrow U(x,y_0)$ and $x \longrightarrow V(x,y_0)$ are constant, and so if $z_1 = x_1 + iy_1$ is any point in D, then

$$\begin{aligned} f(x_1 + iy_0) &= U(x_1, y_0) + iV(x_1, y_0) \\ &= U(x_0, y_0) + iV(x_0, y_0) = f(x_0 + iy_0). \end{aligned}$$

Since $\dfrac{\partial U}{\partial y}(x_1,y) = \dfrac{\partial V}{\partial y}(x_1,y) = 0$, the functions $y \longrightarrow U(x_1,y)$ and $y \longrightarrow V(x_1,y)$ are constant. Thus

$$\begin{aligned} f(x_1 + iy_1) &= U(x_1, y_1) + iV(x_1, y_1) \\ &= U(x_1, y_0) + iV(x_1, y_0) \\ &= f(x_1 + iy_0) = f(x_0 + iy_0). \end{aligned}$$

So $f(z_1) = f(z_0)$, and f is constant on the disc D. ■

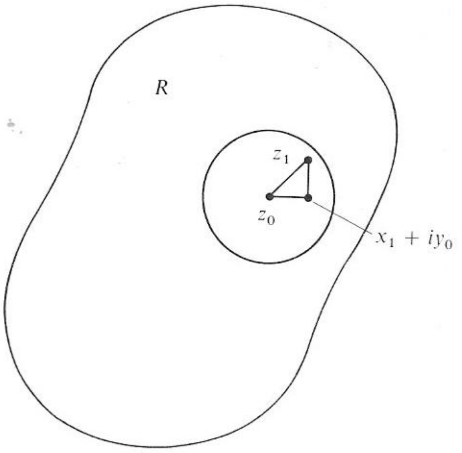


Fig. 23

This theorem can be used to prove some rather more surprising results.

Theorem 17

If f is analytic on a region R and one of the following holds

(i) its real part U is constant on R,

(ii) its imaginary part V is constant on R,

(iii) its modulus is constant on R,

then f is constant on R.

Proof

We prove parts (i) and (iii), leaving part (ii) for you (see Self-Assessment Question 3).

(i) Suppose that $f = U + iV$, and that U is constant on R. Then $\frac{\partial U}{\partial x}(x, y) = \frac{\partial U}{\partial y}(x, y) = 0$ for all x and y such that $x + iy$ is in R. But

$$\begin{aligned} f'(x+iy) &= \frac{\partial U}{\partial x}(x, y) + i\frac{\partial V}{\partial x}(x, y) \\ &= \frac{\partial U}{\partial x}(x, y) - i\frac{\partial U}{\partial y}(x, y), \end{aligned}$$

since the Cauchy–Riemann equations are satisfied. Thus $f' = 0$ on R, and f is constant.

(iii) Suppose that $f = U + iV$, and that $|f|$ is constant on R. Then $|f|^2 = U^2 + V^2$ is constant on R and so the partial derivatives of $U^2 + V^2$ are zero: if $x + iy \in R$,

$$U(x, y)\cdot\frac{\partial U}{\partial x}(x, y) + V(x, y)\cdot\frac{\partial V}{\partial x}(x, y) = 0,$$

and

$$U(x, y)\cdot\frac{\partial U}{\partial y}(x, y) + V(x, y)\cdot\frac{\partial V}{\partial y}(x, y) = 0.$$

Using the Cauchy–Riemann equations, we rewrite these equations in terms of $\frac{\partial U}{\partial x}(x, y)$ and $\frac{\partial U}{\partial y}(x, y)$:

$$U(x, y)\cdot\frac{\partial U}{\partial x}(x, y) - V(x, y)\cdot\frac{\partial U}{\partial y}(x, y) = 0,$$

$$U(x, y)\cdot\frac{\partial U}{\partial y}(x, y) + V(x, y)\cdot\frac{\partial U}{\partial x}(x, y) = 0.$$

Attempting to solve for $\frac{\partial U}{\partial x}(x, y)$ and $\frac{\partial U}{\partial y}(x, y)$, we obtain:

$$\{[U(x, y)]^2 + [V(x, y)]^2\}\cdot\frac{\partial U}{\partial x}(x, y) = 0,$$

$$\{[U(x, y)]^2 + [V(x, y)]^2\}\cdot\frac{\partial U}{\partial y}(x, y) = 0.$$

But $U^2 + V^2$ is constant over R. If it is non-zero $\frac{\partial U}{\partial x}(x, y) = \frac{\partial U}{\partial y}(x, y) = 0$ and by part (i) f is constant. If $U^2 + V^2$ is zero then f is a constant also (namely zero). ■

The Cauchy–Riemann equations are not the only way of proving the theorems we have looked at: they can be proved by complex analytic methods.

If you are unfamiliar with differentiability of functions of two variables, you may omit the rest of this section (as far as the summary on page 148).

In the rest of this section we consider the Cauchy–Riemann equations from a slightly different point of view. A complex function may be thought of as a function from $\mathbf{R}^2$ to $\mathbf{R}^2$, as we explained at the beginning of this section; so complex analysis is a special case of the analysis of functions from $\mathbf{R}^2$ to $\mathbf{R}^2$. We are going to describe a little of that kind of analysis, just enough to point out what makes complex functions special.

Let us consider functions from $\mathbf{R}^2$ to $\mathbf{R}^2$, for the moment quite independent of complex analysis, and ask what it means to say that such a function is differentiable. Suppose that ϕ is a function from some subset of $\mathbf{R}^2$ into $\mathbf{R}^2$. Then $\phi(x, y) \in \mathbf{R}^2$, and so defines an ordered pair of functions (U, V); in other words ϕ defines, and is defined by, a pair of functions U, V of two variables such that $\phi(x, y) = \big(U(x, y), V(x, y)\big)$. Suppose that U and V are differentiable at a point (x_0, y_0).

Then

$$U(x_0 + \xi, y_0 + \eta) = U(x_0, y_0) + \frac{\partial U}{\partial x}(x_0, y_0)\cdot \xi + \frac{\partial U}{\partial y}(x_0, y_0)\cdot \eta + \rho(\xi, \eta)$$

$$V(x_0 + \xi, y_0 + \eta) = V(x_0, y_0) + \frac{\partial V}{\partial x}(x_0, y_0)\cdot \xi + \frac{\partial V}{\partial y}(x_0, y_0)\cdot \eta + \sigma(\xi, \eta).$$

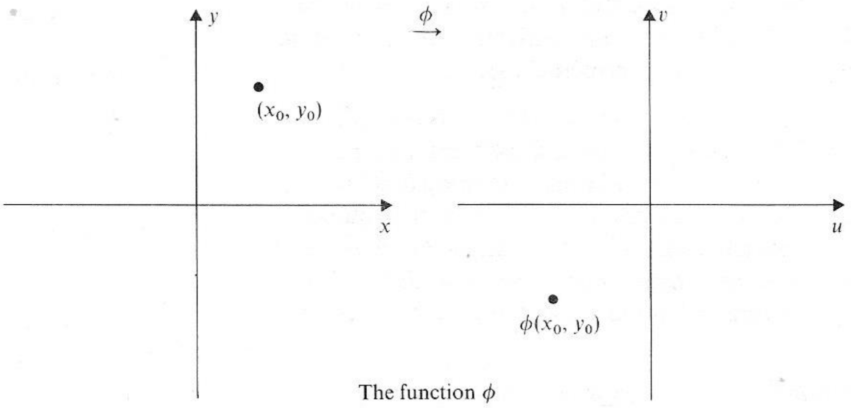
The terms involving the partial derivatives look suspiciously like the result of multiplying two matrices. If we write elements of $\mathbf{R}^2$ as column matrices we may combine the two equations above into one matrix equation to give

$$\phi(x_0 + \xi, y_0 + \eta) = \phi(x_0, y_0) + A\begin{bmatrix}\xi\\ \eta\end{bmatrix} + \begin{bmatrix}\rho(\xi, \eta)\\ \sigma(\xi, \eta)\end{bmatrix},$$

where $\phi(x, y)$ now represents the matrix $\begin{bmatrix}U(x, y)\\ V(x, y)\end{bmatrix}$, and A is the matrix of partial derivatives

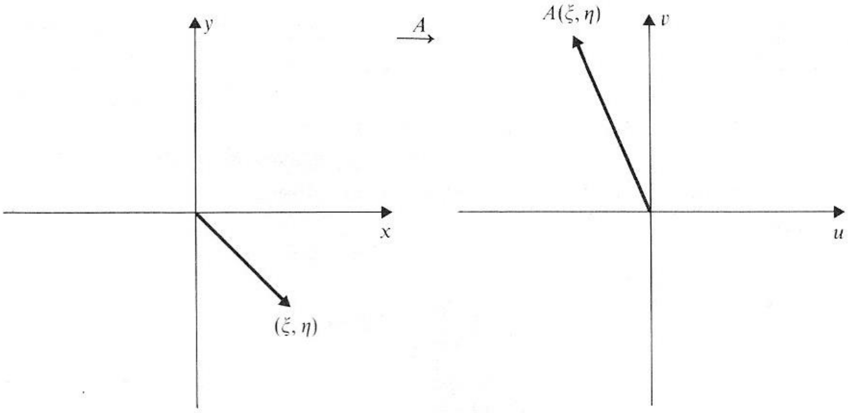
$$\begin{bmatrix}\dfrac{\partial U}{\partial x}(x_0, y_0) & \dfrac{\partial U}{\partial y}(x_0, y_0)\\[2ex] \dfrac{\partial V}{\partial x}(x_0, y_0) & \dfrac{\partial V}{\partial y}(x_0, y_0)\end{bmatrix}.$$

The "remainder" term $\begin{bmatrix}\rho(\xi, \eta)\\ \sigma(\xi, \eta)\end{bmatrix}$ obeys a certain limit condition which we gave explicitly when discussing the Cauchy–Riemann equations. It is enough to say here that if (ξ, η) is close to $(0, 0)$ then the remainder is small. Thus our matrix equation says that if U and V are differentiable at (x_0, y_0) then, for points near (x_0, y_0), ϕ can be approximated by a linear transformation. In other words, if we think of $(x_0 + \xi, y_0 + \eta)$ as determining an arrow with blunt end at (x_0, y_0) which corresponds to the geometric vector represented by $\begin{bmatrix}\xi\\ \eta\end{bmatrix}$, then the image of $(x_0 + \xi, y_0 + \eta)$ is the arrow with blunt end at $\phi(x_0, y_0)$ which corresponds to the geometric vector represented by $A\begin{bmatrix}\xi\\ \eta\end{bmatrix}$. We picture the situation in a sequence of figures.

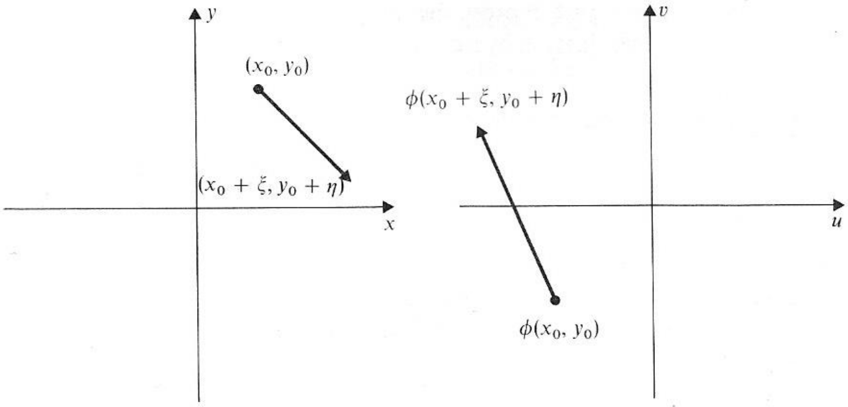

y
$\xrightarrow{\phi}$
v
(x_0, y_0)
x
u
$\phi(x_0, y_0)$

The function ϕ

Fig. 24


y
$\xrightarrow{A}$
$A(\xi, \eta)$
v
x
u
(ξ, η)

The linear transformation represented by A
Fig. 25


y
v
(x_0, y_0)
$\phi(x_0 + \xi, y_0 + \eta)$
$(x_0 + \xi, y_0 + \eta)$
x
u
$\phi(x_0, y_0)$

The approximate effect of ϕ on points near (x_0, y_0)
Fig. 26

Our definition of differentiability is based on the results we have just described. A function $\phi : \mathbf{R}^2 \longrightarrow \mathbf{R}^2$ is differentiable at a point (x_0, y_0) if it can be approximated by a linear transformation for points near (x_0, y_0). To be more explicit, ϕ is differentiable at (x_0, y_0) if there is a matrix A such that

$$\phi(x_0 + \xi, y_0 + \eta) = \phi(x_0, y_0) + A\begin{bmatrix}\xi \\ \eta\end{bmatrix} + \text{remainder term},$$

where the remainder term satisfies a certain limit condition. It can be shown that if ϕ is differentiable then A is unique, the x and y components of ϕ have partial derivatives, and A is just the matrix of partial derivatives.

This may look rather different from the definition of differentiability that you are used to from real analysis. But remember that the derivative of a function of a real variable may be thought of as a scale factor, and that a linear transformation in a one-dimensional space must be multiplication by a constant, and you will see that the two definitions are in principle identical.

Let us now return to complex analysis. The point of our discussion is that it enables us to view the Cauchy–Riemann equations in a different, and perhaps more revealing, light. Evidently it is not enough to say of a mapping of $\mathbf{R}^2$ into $\mathbf{R}^2$ that it is differentiable (in the sense we have just explained) for that mapping to define a function of a complex variable which is differentiable in the sense of complex analysis. The transformation which approximates a differentiable function of a complex variable is multiplication by a complex number. (See Theorem 8 of Section 3.2.)

Multiplication by a complex number, considered as a transformation of $\mathbf{R}^2$, is indeed linear: if $\begin{bmatrix} \xi \\ \eta \end{bmatrix}$ corresponds to $\xi + i\eta$, then the matrix corresponding to $(\alpha + i\beta)(\xi + i\eta) = (\alpha\xi - \beta\eta) + i(\beta\xi + \alpha\eta)$ is

$$\begin{bmatrix} \alpha\xi - \beta\eta \\ \beta\xi + \alpha\eta \end{bmatrix} = \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix}.$$

So the matrix of partial derivatives of a mapping from $\mathbf{R}^2$ to $\mathbf{R}^2$ defined by a function of a complex variable which is differentiable at $z_0 = x_0 + iy_0$ must have equal terms on the main diagonal and terms differing in sign on the other diagonal. The matrix of partial derivatives is

$$\begin{bmatrix} \dfrac{\partial U}{\partial x}(x_0, y_0) & \dfrac{\partial U}{\partial y}(x_0, y_0) \\ \dfrac{\partial V}{\partial x}(x_0, y_0) & \dfrac{\partial V}{\partial y}(x_0, y_0) \end{bmatrix}$$

and this means that $\dfrac{\partial U}{\partial x}(x_0, y_0) = \dfrac{\partial V}{\partial y}(x_0, y_0)$ and $\dfrac{\partial V}{\partial x}(x_0, y_0) = -\dfrac{\partial U}{\partial y}(x_0, y_0)$. In other words, the Cauchy–Riemann equations just express the fact that the matrix of partial derivatives corresponds to multiplication by a complex number.

Summary

You should by now be beginning to realize that the condition of differentiability for a complex function is rather a subtle one. The first suggestion we had that this is so occurred in Section 3.1, with the result that functions like the real part and the modulus functions are not differentiable. The section you have just been reading contains further evidence, in the form of the Cauchy–Riemann equations, and the special position of complex functions in the set of functions from $\mathbf{R}^2$ to $\mathbf{R}^2$.

Also, the Cauchy–Riemann equations are a useful tool in determining whether a function is differentiable.

Self-Assessment Questions

1. Verify that the real and imaginary parts of the complex function $f(z) = 1/z$, $z \neq 0$, satisfy the Cauchy–Riemann equations. What can you deduce about the function f?
2. Show that the function $f(z) = e^{-x}(\cos y - i \sin y)$ is entire, by using the Cauchy–Riemann equations.
3. Show by using the Cauchy–Riemann equations that if f is analytic on a region R and its imaginary part V is constant then f is constant.

Solutions

1. Since $f(z) = \dfrac{1}{z} = \dfrac{1}{x + iy} = \dfrac{x - iy}{x^2 + y^2}$, we have

$$U(x, y) = \frac{x}{x^2 + y^2},\ V(x, y) = -\frac{y}{x\cancel{y}^2 + y^2},\ (x, y) \neq (0, 0).$$

The partial derivatives of U and V are

$$\frac{\partial U}{\partial x}(x, y) = \frac{y^2 - x^2}{(x^2 + y^2)^2}, \qquad \frac{\partial U}{\partial y}(x, y) = -\frac{2xy}{(x^2 + y^2)^2},$$

$$\frac{\partial V}{\partial x}(x, y) = \frac{2xy}{(x^2 + y^2)^2}, \qquad \frac{\partial V}{\partial y}(x, y) = \frac{y^2 - x^2}{(x^2 + y^2)^2}.$$

Thus $\dfrac{\partial U}{\partial x}(x, y) = \dfrac{\partial V}{\partial y}(x, y)$ and $\dfrac{\partial V}{\partial x}(x, y) = -\dfrac{\partial U}{\partial y}(x, y)$, when $(x, y) \neq (0, 0)$.

Since the partial derivatives are continuous provided $(x, y) \neq (0, 0)$, and the Cauchy–Riemann equations are satisfied, the function f is analytic on the region $\{z : z \neq 0\}$.

2. Since $f(z) = e^{-x}(\cos y - i \sin y)$, we have

$$U(x, y) = e^{-x} \cos y, \qquad V(x, y) = -e^{-x} \sin y.$$

The partial derivatives of U and V are

$$\frac{\partial U}{\partial x}(x, y) = -e^{-x} \cos y, \qquad \frac{\partial U}{\partial y}(x, y) = -e^{-x} \sin y,$$

$$\frac{\partial V}{\partial x}(x, y) = e^{-x} \sin y, \qquad \frac{\partial V}{\partial y}(x, y) = -e^{-x} \cos y.$$

These functions are continuous and the Cauchy–Riemann equations are satisfied for all (x, y). Thus f is entire. (In fact, $f(z) = e^{-z}$.)

3. Suppose that $f = U + iV$ and V is constant on R.

Then $\dfrac{\partial V}{\partial x}(x, y) = \dfrac{\partial V}{\partial y}(x, y) = 0$ for all (x, y) such that $x + iy \in R$. By the Cauchy–Riemann equations

$$\begin{aligned} f'(x + iy) &= \frac{\partial U}{\partial x}(x, y) + i\frac{\partial V}{\partial x}(x, y) \\ &= \frac{\partial V}{\partial y}(x, y) + i\frac{\partial V}{\partial x}(x, y) = 0, \end{aligned}$$

so f is constant on R.

3.9 PROBLEMS

1. Show that cos and sin are entire, and satisfy $\cos' = -\sin$ and $\sin' = \cos$, by using the Cauchy–Riemann equations and the formulas

$$\cos(x + iy) = \cos x \cosh y - i \sin x \sinh y$$

$$\sin(x + iy) = \sin x \cosh y + i \cos x \sinh y.$$

(The functions on the right-hand sides of these equations are the usual functions of real analysis.)

2. Show that if a function f is real-valued and analytic on a region R then it is constant on R.

3. This problem is optional since it demands some knowledge of differentiability of functions of two variables.

Show that if we define a function f by

$$f(z) = \begin{cases} \sqrt{xy}, & \text{if } xy \geqslant 0 \\ i\sqrt{-xy}, & \text{if } xy < 0 \end{cases} \text{(where } z = x + iy\text{)},$$

then the real and imaginary parts of f satisfy the Cauchy–Riemann equations at $(0, 0)$, but f is not differentiable at 0. (This example shows that in Theorem 14 (of Section 3.8) some extra condition, other than just that the real and imaginary parts of a function satisfy the Cauchy–Riemann equations is essential.)

Solutions

1. The real and imaginary parts of cos are, respectively,

$$U(x, y) = \cos x \cosh y \quad \text{and} \quad V(x, y) = -\sin x \sinh y.$$

The partial derivatives are

$$\frac{\partial U}{\partial x}(x, y) = -\sin x \cosh y, \qquad \frac{\partial U}{\partial y}(x, y) = \cos x \sinh y,$$

$$\frac{\partial V}{\partial x}(x, y) = -\cos x \sinh y, \qquad \frac{\partial V}{\partial y}(x, y) = -\sin x \cosh y.$$

These functions are continuous, the Cauchy–Riemann equations are satisfied, so cos is differentiable everywhere; in other words, cos is entire. We have

$$\begin{aligned} \cos'(x + iy) &= \frac{\partial U}{\partial x}(x, y) + i\frac{\partial V}{\partial x}(x, y) \\ &= -\sin x \cosh y - i \cos x \sinh y \\ &= -\sin(x + iy). \end{aligned}$$

The procedure for sin is similar.

2. If f is real-valued on R, then the imaginary part of f is a constant (namely zero); so by Self-Assessment Question 3 of Section 3.8, f, which is analytic on R, is constant on R.

3. The real and imaginary parts of f are

$$U(x, y) = \begin{cases} \sqrt{xy}, & \text{if } xy \geqslant 0 \\ 0, & \text{if } xy < 0, \end{cases} \qquad V(x, y) = \begin{cases} 0, & \text{if } xy \geqslant 0 \\ \sqrt{-xy}, & \text{if } xy < 0. \end{cases}$$

We calculate the partial derivatives of U and V at $(0, 0)$:

$$\frac{U(\alpha, 0) - U(0, 0)}{\alpha} = 0 \qquad \text{for all } \alpha \neq 0,$$

so $\dfrac{\partial U}{\partial x}(0, 0) = 0$. Similarly, $\dfrac{\partial U}{\partial y}(0, 0) = \dfrac{\partial V}{\partial x}(0, 0) = \dfrac{\partial V}{\partial y}(0, 0) = 0$.

The Cauchy–Riemann equations are satisfied at $(0, 0)$. Consider now what happens to $\dfrac{f(h) - f(0)}{h}$ if h is not a real or imaginary number. Suppose for example $\operatorname{Re} h = \operatorname{Im} h$, so that $h = \alpha + i\alpha$ for some non-zero real number α. Then

$$\frac{f(h) - f(0)}{h} = \frac{\sqrt{\alpha^2}}{\alpha + i\alpha} = \frac{|\alpha|}{\alpha} \cdot \frac{1}{1 + i}.$$

Now any disc centred at 0 contains numbers of the form $\alpha + i\alpha$ with $\alpha > 0$ and with $\alpha < 0$, so it contains numbers h for which

$$\frac{f(h) - f(0)}{h} = \frac{1}{1 + i}$$

and also numbers h for which

$$\frac{f(h) - f(0)}{h} = -\frac{1}{1 + i}$$

(and also numbers h for which $[f(h) - f(0)]/h = 0$—namely real and imaginary numbers). It follows that f cannot be differentiable at 0.

It is easy to convince oneself that the partial derivatives of U and V are not continuous at $(0, 0)$.

NOTATION

The following items of notation are explained on the pages given. Most other items of notation used in the text may be found in the *Mathematical Handbook* for M100, The Mathematics Foundation Course, or in the *Handbook* for M231, Analysis.

$\mathbf{R}$	8
$\mathbf{R} \times \mathbf{R}$	8
$x + iy$	8
$\mathbf{C}$	9
$+$ (for $\mathbf{C}$)	9
$\times$ (for $\mathbf{C}$)	9
i	10
$\mathrm{Re}(a + ib)$, $\mathrm{Re}\, z$	11
$\mathrm{Im}(a + ib)$, $\mathrm{Im}\, z$	11
$z = x + iy$	11
$0\ (\in \mathbf{C})$	12
$1\ (\in \mathbf{C})$	12
$-$(for $\mathbf{C}$)	12
$(\mathbf{C}, +, \times)$	17
$(\mathbf{C}, +)$, $(\mathbf{C}. \times)$	17
$r(\cos\theta + i\sin\theta)$	21
$\lvert z\rvert = \lvert x + iy\rvert$	21
$\mathrm{Arg}\, z$	21
$\div$ (for $\mathbf{C}$)	22
$\bar{z}$	24
$d:(z, w) \longrightarrow \lvert z - w\rvert, (z,w \in \mathbf{C} \times \mathbf{C})$	36
$\mathbf{N}$	37
$\{z_n\}$	37
$\mathbf{Q}$	38
$\bigcap_{i\in\mathbf{N}} S_i$	39
$[z_1, z_2]$	44
$f + g$	51
$f \cdot g$	51
f/g	51
$f \circ g$	51
$\lvert f\rvert$	52
$\mathrm{Re} f$	52
$\mathrm{Im} f$	52
$\bar{f}$	52
$f(z) = U(x, y) + iV(x, y)$	52
$f(S)$	53
$f^{-1}(S)$	54
f^{-1}	55
$\lim_{z\to a} f(z)$	59
$\mathbf{C} - A$	64
$\bar{S}$	70
$[z, z_1, z_2, \ldots, z_{n-1}, z']$	80
$\max\{f(z): z \in S\}$	92
$\min\{f(z): z \in S\}$	93
$d_A(z)$	93
$f'(a)$	102
f'	102
$\dfrac{d}{dz}$	104
γ	111
$f(\gamma)$	111
exp	120
$\exp z$	120
e^z	120
cos	123
$\cos z$	123
sin	123
$\sin z$	123
cosh	125
$\cosh z$	125
sinh	125
$\sinh z$	125
tanh	127
$\tanh z$	127
Log	130
$\mathrm{Log}\, z$	130
L_R	131
L	133
z^α	135
$\dfrac{\partial F}{\partial x}(\alpha, \beta), \dfrac{\partial F}{\partial y}(\alpha, \beta)$	139

INDEX

Accumulation point, 43
Addition
  of complex numbers, Cartesian form, 9
  of complex numbers, polar form, 21
  of functions, 51
Analytic on R, 102
Annulus
  closed, 37
  open, 37
Argument of a complex number, 21
  principal value of, 21
Axis
  imaginary, 20
  real, 20

Bolzano-Weierstrass Theorem, 43
Boundary
  of a set, 72
  point of a set, 72
Bounded
  function on a set, 88
  set, 37, 90
Boundedness Theorem, 92
Branch
  of the αth power, 134
  of the logarithm associated with a fundamental region of exp, 131
  of the square root, 134

Cartesian coordinates, 20
Cartesian equation, 42
Cauchy-Riemann equations, 140
Chain Rule, 103
Closed
  annulus, 37
  disc, 37
  rectangle, 38
  set, 70
Closure of a set, 70
Cluster point, 43, 69
Compact set, 89
Complement of a set, 64
Complex function, 50
Complex number
  conjugate of, 24
  imaginary part of, 11
  modulus of, 21
  real part of, 11
Complex number system, 9
Complex plane, 20
Composition of functions, 51
Conjugate, 24
Conjugation function, 52
Connected set, 77
Constant function, 51
Continuous
  at a, 61, 73
  at (α, β), 139
  function, 61, 73
  uniformly, 95
Convergent sequence of complex numbers, 37
Convex set, 79
Coordinate lines, 112
cos, 123
cosh, 125
Covering property, 89
Curve, 111
Cut plane, 86

Degree of a polynomial, 16
de Moivre's Theorem, 30
Derivative, 102
  at a, 102
  at (α, β), partial, 139
Derived function, 102
Difference of two complex numbers, 12
Differentiable
  at a, 102
  at (α, β), 139
Disc, 36
  closed, 37
  open, 37
  punctured, 62
  unit, 37
Disconnected set, 77
Disjoint sets, 77
Disjoint union of two sets, 77
Distance
  between two complex numbers, 36
  from a point to a set, 93
  from (α, β) to (x, y), 139
  function, 36
Divisible polynomial, 16
Division
  of complex numbers, Cartesian form, 22
  of complex numbers, polar form, 23
  of functions, 51
Domain, 50

Entire function, 102
Equal complex numbers, 10
Equating real and imaginary parts, 11
Even function, 56
Exponential function, 120

Field, 13
  ordered, 12
  properties of complex numbers, 12
Function, 50
  bounded on a set, 88
  complex, 50
  conjugation, 52
  constant, 51
  continuous, 61
  continuous at a, 61
  continuous at (α, β), 139
  derived, 102
  entire, 102
  exponential, 120
  hyperbolic, 125
  identity, 51
  inverse of, 55
  locally bounded at a point on a set, 88
  modulus, 36, 52
  of a real variable, 52
  operations on, 51
  periodic, 121
  power, 134
  polynomial, 51
  rational, 51
  real, 52
  real-valued, 52
  trigonometric, 122
  value of at z, 50
Fundamental region for exp, 131
Fundamental Theorem of Algebra, 17

Half plane, 33, 64
  open, 63
Hyperbolic functions, 122

Identity function, 51
Image of a set under a function, 53
Imaginary axis, 20
Imaginary part of a complex number, 11
Instantaneous
  speed, 111
  velocity, 111
Interior point of a set, 62
Intersection of a collection of sets, 38
Intervals, 72
Inverse image of a set under a function, 54
Inverse of a function, 55
Isolated point of a set, 69
Isometry, 57

Leibniz notation, 104
Limit near a, 59, 73
Limit of sequence, 37
Line segment, 42
Locally
  bounded function at a point on a set, 88
  constant function at a point, 86
Logarithm
  branch associated with a fundamental region of exp, 131
  principal branch of, 130

Matrix of partial derivatives, 146
Maximum of a function on a set, 92
Metric space, 46
Minimum of a function on a set, 92
Modulus
  function, 36, 52
  of a complex number, 21
Multiplication
  of complex numbers, Cartesian form, 9
  of complex numbers, polar form, 22
  of functions, 51

Neighbourhood of a point, 63
Nested Rectangles Theorem, 39
Nested sequence of sets, 37
nth roots of a complex number, 25

One-one function, 55
Odd function, 56
Open
  annulus, 37
  disc, 37
  rectangle, 38
  set, 63
Ordered field, 13

Partial derivatives at (α, β), 139
Period, 121
Periodic function, 121
Polar coordinates, 20
Polar form of a complex number, 21
Polygonal line from z to z', 80
Polygonally
  connected set, 80
  connected to, 80

COMPLEX ANALYSIS

1 Complex Numbers
2 Continuous Functions
3 Differentiation
4 Integration
5 Cauchy's Theorem I
6 Taylor Series
7 NO TEXT
8 Singularities
9 Cauchy's Theorem II
10 The Calculus of Residues
11 Analytic Functions
12 NO TEXT
13 Properties of Analytic Functions
14 Laplace Transforms
15 Number Theory
16 Boundary Value Problems

Course Team

Chairman:	Dr. G. A. Read	Senior Lecturer
Authors:	Dr. P. D. Bacsich	Lecturer
	Dr. M. Crampin	Lecturer
	Mr. N. W. Gowar	Senior Lecturer
	Dr. R. J. Wilson	Lecturer
Editor:	Mr. R. J. Knight	Lecturer
B.B.C.:	Mr. D. Saunders	

With assistance from:

Mr. D. Goldrei	Course Assistant
Mr. H. Hoggan	B.B.C.
Mr. R. J. Margolis	Staff Tutor
Dr. A. R. Meetham	Staff Tutor
Mr. J. E. Phythian	Staff Tutor
Mrs. P. M. Shepheard Rogers	Course Assistant
Dr. C. A. Rowley	Course Assistant

0 335 05550 8